The Dissertation Committee for Junhua Zhao Certifies that this is the approved version of the following dissertation:

Polar localization of a group II intron-encoded reverse transcriptase and its effect on retrohoming site distribution in the *E. coli* genome

Committee:

Alan M. Lambowitz, Supervisor

George Georgiou

Makkuni Jayaram

Richard J. Meyer

James R. Walker

Polar localization of a group II intron-encoded reverse transcriptase and its effect on retrohoming site distribution in the *E. coli* genome

by

Junhua Zhao, B.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August, 2007

Dedication

To my parents, husband and the whole family

Acknowledgements

I would especially like to thank my supervisor, Dr. Alan Lambowitz, for his thoughtful inspirations, academic guidance, generous support, and patience through these years. He is the most hardworking and persistent mentor in my research career. I am also so grateful to the members of my dissertation committee, Drs. George Georgiou, Makkuni Jayaram, Richard Meyer, and James Walker, for their valuable advice and great support.

I thank all the members of the laboratory for their friendship and helpful suggestions, especially Drs. Sabine Mohr, Roland Saldanha, Joseph San Filippo, Xiaoxia Cui, Georg Mohr, Jin Zhong, Manabu Matsuura, Jiri Perutka, Huatao Guo and Yue Jiang. I would also like to thank Mr. Jun Yao for scientific discussions and help with experimental equipments and supplies.

Finally, I would like to thank my parents for the huge support during my graduate career, and thank my husband, Hui, for his love and technical support in data storage and analysis.

Polar localization of a group II intron-encoded reverse transcriptase and its effect on retrohoming site distribution in the *E. coli* genome

Publication No._____

Junhua Zhao, Ph.D.

The University of Texas at Austin, 2007

Supervisor: Alan M. Lambowitz

The *Lactococcus lactis* Ll.LtrB group II intron encodes a reverse transcriptase (LtrA protein), which binds the intron RNA to promote RNA splicing and intron mobility. Mobility occurs by intron RNA reverse splicing directly into a DNA strand and reverse transcription by LtrA. I used LtrA-GFP fusions and immunofluorescence microscopy to show that LtrA localizes to the cellular poles in both *Escherichia coli* and *L. lactis*. This polar localization occurs with or without co-expression of the intron RNA, is observed over a wide range of cellular growth rates and expression levels, and is independent of replication origin function. The same localization pattern was found for three non-overlapping LtrA subsegments, reflecting dependence on common redundant signals and/or protein physiochemical properties. When coexpressed in *E. coli*, LtrA interferes with the polar localization of the *Shigella* IcsA protein, which mediates polarized actin tail assembly, suggesting competition for a common localization determinant.

In *E. coli*, the Ll.LtrB intron inserts preferentially into the chromosomal *ori* and *ter* regions, which are pole localized during much of the cell cycle. Thus, the polar localization of LtrA could account for the preferential insertion of the Ll.LtrB intron in these regions. I established a high throughput method using cellular array and automated fluorescence microscopy for screening transposon-induced mutants, and identified five *E. coli* genes (*gppA, uhpT, wcaK, ynbC,* and *zntR*) in which disruptions result in increased proportion of cells having diffuse LtrA distribution. This altered localization is correlated with a more uniform distribution of Ll.LtrB insertion sites throughout the *E. coli* genome. Finally, I find that altered LtrA localization in all five disruptants is correlated with accumulation and more diffuse intracellular distribution of polyphosphate, and that a *ppx* disruptant, which also results in polyphosphate accumulation, shows similar LtrA mislocalization. These findings may reflect interaction between LtrA and intracellular polyphosphate. My findings support the hypothesis that the intracellular localization of LtrA is a major determinant of Ll.LtrB insertion site preference in the *E. coli* genome. Further, they show that alterations in polyphosphate metabolism can lead to protein mislocalization, and suggest that polyphosphate is an important factor affecting intracellular protein localization.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Group II introns are mobile retroelements that are found in bacteria and organelles (reviewed in Lambowitz & Zimmerly, 2004). The autocatalyic intron RNA and its multifunctional intron-encoded protein (IEP) together form a ribonucleoprotein (RNP) that carries out RNA splicing and intron mobility (Figure 1.1, Belfort *et al.,* 2002; Lambowitz & Zimmerly, 2004). The IEP promotes splicing by stabilizing the catalytically active structure of the intron RNA and then remains bound to the excised intron lariat RNA in RNPs. RNPs mediate intron mobility with the intron RNA first inserting (reverse-splicing) into one strand of the DNA substrate and then being reverse transcribed by the IEP. Mobile group II introns use this mechanism to insert into specific DNA target sites at high frequency in a process called retrohoming and into ectopic sites that resemble the normal homing site at low frequency in a process called retrotransposition (Lambowitz & Zimmerly, 2004). The latter process enabled mobile group II introns to become widely dispersed among bacterial species and may have been used to invade eukaryotic nuclear genomes, where mobile group II introns are thought to have evolved into both spliceosomal introns (Sharp, 1991) and non-LTR-retrotransposons (Eickbush, 1994; Zimmerly *et al.,* 1995b). However, while the major biochemical steps in their DNA integration mechanism have been elucidated, little is known about how mobile group II introns function in the context of different cellular structures and compartments or how their mobility is coordinated with other cellular processes, such as cell division or DNA replication.

1.1 STRCTURE OF *LACTOCOCCUS LACTIS* GROUP II INTRON LL.LTRB INTRON RNA

The *Lactococcus lactis* Group II intron Ll.LtrB has been an important model system for studying mobile group II introns, because it expresses and splices efficiently in *E. coli* (Matsuura *et al.,* 1997). This intron was discovered in a relaxase gene (*ltrB*) in an *L. lactis* conjugative element pRS01, where its splicing is essential to produce the functional relaxase for conjugation (Mills *et al.,* 1996, 1997).

Like other group II introns, Ll.LtrB RNA folds into the conserved secondary structure, consisting of typical six double-helical domains, domains I-VI (DI-DVI, respectively; Figure 1.2). DI contains exon-binding sites 1 and 2 (EBS1 and EBS2, respectively) and δ, which base pair with intron-binding sites 1 and 2 (IBS1 and IBS2, respectively) and δ' in the flanking 5'- and 3'- exon (Mills *et al.,* 1996; Mohr *et al.,* 2000). These base-pairing interactions position the 5'- and 3'- splice sites at the intron's active site and are also used for DNA target site recognition during intron mobility (Lambowitz & Zimmerly, 2004). DII and DIII contribute to RNA folding and the formation of the catalytic core (Fedorova *et al.,* 2003). DIV encodes the open reading frame (ORF) for the IEP (in Ll.LtrB, the IEP is denoted LtrA protein), but this domain is not required for ribozyme activity. DV is the most conserved and interacts with DI to form the minimal catalytic core. DVI contains the branch-point nucleotide residue, generally a bulged A residue (Michel & Ferat, 1995; Mills *et al.,* 1996; Lambowitz & Zimmerly, 2004; Pyle & Lambowitz, 2006).

The conserved group II intron domains interact with each other via tertiary contacts to form a conserved three-dimensional structure that is required for the RNA's catalytic activity (reviewed in Pyle & Lambowitz, 2006). This folded RNA structure contains an active site that binds the 5'- and 3'- splice sites and the branch-point nucleotide residue and uses specifically bound Mg^{2+} ions to activate the appropriate

phosphodiester bonds for catalysis (Pyle & Lambowitz, 2006). Splicing occurs by two sequential transfesterification reactions (Figure 1.3, Lambowitz & Zimmerly, 2004) that result in the formation of an intron lariat RNA, analogous to the splicing mechanism of spliceosomal introns in higher organisms (Michel & Ferat, 1995).

1.2 CHARACTERISTICS OF LL.LTRB INTRON ENCODED PROTEIN (IEP)-LTRA

Like other group II intron IEPs, the LtrA protein encoded by the Ll.LtrB intron is a multifunctional protein with four conserved domains: reverse transcriptase (RT), corresponding to the fingers and palm regions of retroviral RTs; X, associated with maturase activity, corresponding to the RT thumb; DNA-binding (D); and DNA endonuclease (En) (Figure 1.4). The RT and X domains function together to bind the intron RNA and stabilize its active structure for RNA splicing and reverse splicing (Cui *et al.,* 2004; Blocker *et al.,* 2005). Domain D is required for efficient reverse splicing into double-stranded DNA, while En domain cleaves the opposite strand to generate the primer for reverse transcription of the intron RNA during intron mobility (San Filippo & Lambowitz, 2002). Although En-dependent retrohoming is favored, when En cleavage is blocked by mutation, Ll.LtrB can still retrohome by using nascent strands at DNA replication forks to prime reverse transcription (Zhong & Lambowitz, 2003; Yao *et al.,* 2005). Analogous En-independent mechanisms are also used for retrotransposition of the Ll.LtrB intron to ectopic sites (Ichiyanagi *et al.,* 2002; Coros *et al.,* 2005).

1.3 LL.LTRB INTRON MOBILITY IN *E. COLI*

Ll.LtrB RNPs initiate mobility by recognizing relatively long (30-35 bp) DNA target sites by a combination of IEP interactions and base pairing of the intron RNA to the DNA target sequence (Singh & Lambowitz, 2001). The IEP first recognizes a small number of specific nucleotide residues in the distal 5'- exon region of the DNA target

site, including T-23, G-21, and A-20, via major groove interactions. These base interactions along with phosphate backbone interactions on one face of the helix trigger the unwinding of the double-strand target DNA, enabling the intron RNA to base-pair to the target DNA at positions –12 to +3 for reverse splicing of the intron in the top strand (Singh & Lambowitz, 2001). Second-strand cleavage occurs after reverse splicing and requires additional interactions between the IEP and the 3' exon, most significantly recognition of T+5 (San Filippo & Lambowitz, 2002). The DNA endonuclease (En) domain in the IEP cleaves the second-strand and uses the 3' end at the cleavage site as a primer for reverse transcription of the inserted intron RNA.

The downstream steps involved in the integration of the intron cDNA into the host genome use a DNA repair mechanism independent of RecA function (Cousineau *et al.,* 1998; Smith *et al.,* 2005). Smith *et al.* (2005) suggested that after cDNA synthesis by the intron IEP, RNase H may play a role in degradation of the intron RNA template and the 5' to 3' exonuclease activity of polymerase I likely removes RNA primers, prior to second-strand DNA synthesis by the host replicative polymerase III, various exonuclease, such as RecJ or MutD, may resect overhangs and finally DNA ligase seals the nicks in DNA (Figure 1.1).

The region of the DNA target site recognized by intron RNA base pairing extends from positions -12 to +3 (relative to the intron-insertion site) and consists of three sequence elements, IBS1 and IBS2 in the 5' exon and δ' in the 3' exon; the complementary intron RNA sequences are EBS1, 2 and δ, while the IEP recognizes just a few specific nucleotide residues in the distal regions and facilitates local DNA melting, enabling the intron RNA to base pair to the IBS1 and δ' sequences (Singh & Lambowitz, 2001). Because most of the DNA target site is recognized by base pairing of the intron RNA, it is possible to retarget the Ll.LtrB intron to insert into desired DNA sites simply

by modifying the intron RNA. This feature enabled the development of the Ll.LtrB intron into a new type of gene targeting vector ("targetron") with programmable DNA target specificity (reviewed in Lambowitz *et al.,* 2005).

Guo *et al*. (2000) developed a plasmid-based intron mobility assay, which I used extensively in my research (Figure 1.5). The intron-donor plasmid in this assay is pACD2X, and the intron-recipient plasmid is pBRR3-ltrB. pACD2X uses a T7*lac* promoter to express a 0.9-kb Ll.LtrB-ΔORF intron with short flanking exons, and the intron contains an additional T7 promoter in DIV near its 3' end. The LtrA ORF is expressed from a position just downstream of the 3' exon. The LtrA protein expressed from this downstream *cis* position binds to the intron to promote RNA splicing and then remains tightly bound to the excised intron RNA in RNPs that promote intron mobility. The recipient plasmid contains the wild-type Ll.LtrB target site (ligated Exon 1 (E1) and Exon 2 (E2) sequence of the *ltrB* gene) cloned upstream of a promoterless tet^R gene. The target site is flanked by an upstream *E. coli rrnB* T1 transcription terminator, which terminates both the *E. coli* and T7 RNA polymerases, and a downstream *rrnB* T2 terminator, which terminates *E. coli* but not T7 RNA polymerase (Figure 1.5, Guo *et al.,* 2000). Insertion of the intron carrying the T7 promoter into the target sites activates the expression of the tet^R gene, and mobility frequencies are measured as the ratio of $(Tet^R+Amp^R)/Amp^R$ colonies. This plasmid-based assay was used to determine the detailed target-site recognition rules for the *L. lactis* Ll.LtrB intron and provides a convenient method for analyzing intron insertion efficiency (Guo *et al.,* 2000; Perutka *et al.,* 2004).

1.4 THE LL.LTRB INTRON INSERTS PREFERENTIALLY INTO THE *ORI* REGION OF *E. COLI* GENOME

An Ll.LtrB intron with randomized EBS2, EBS1, and δ sequences inserts at sites distributed throughout the *E. coli* genome, and was used to obtain a collection of gene disruptants, analogous to global transposons mutagenesis (Zhong *et al.*, 2003) The selection of intron-insertion events was facilitated by the introduction of a Tp^R-RAM (retrotransposition-activated selectable marker) in intron DIV (Figure 1.6A, Zhong *et al.,* 2003). In this RAM marker, the tp^R gene is in the reverse orientation to the Ll.LtrB intron, and is disrupted by a group I *td* intron in the same orientation as the Ll.LtrB. After transcription, the *td* intron splices, leaving an intact tp^R gene in Ll.LtrB intron. The integration of the Ll.LtrB intron into *E. coli* genome inserts the tp^R gene, resulting in trimethoprim resistance.

Surprisingly, Zhong *et al.* (2003) found that although the Ll.LtrB intron could insert in any region of the *E. coli* chromosome, the intron-insertions sites obtained from the Ll.LtrB intron library with randomized EBS and δ sequences were strongly clustered around the bidirectional replication origin (*oriC*), with 57% of the sites clustered within 5% of the genome on either side of *oriC* (Figure 1.6B, Zhong *et al.,* 2003). Studies of retrotransposition of wild-type Ll.LtrB intron into ectopic sites in *E. coli* revealed a similar clustering of insertion sites, but in this case, in both the origin (*ori*) and terminus (*ter*) regions (Coros *et al.,* 2005). Because both the *ori* and *ter* regions localize to cell poles after DNA replication (Niki & Hiraga, 1998), the clustering of intron-insertion sites in these regions might reflect the intracellular localization of Ll.LtrB RNPs and/or the contribution of intracellular localized host factors to intron integration. For example, in non-replicating bacterial cells, the major replicative DNA polymerase Pol III in *E. coli*, is localized at cell poles (Onogi *et al.*, 2002), and Pol III was recently suggested to be

responsible for second-strand synthesis during retrohoming and retrotransposition (Figure 1.1, Smith *et al.,* 2005).

1.5 OVERVIEW OF POLAR LOCALIZATION IN BACTERIA

The *ori* and *ter* regions of *E. coli* chromosome are localized near the cellular poles during much of the cell cycle (Niki & Hiraga, 1998; Niki *et al.,* 2000; Draper & Gober, 2002). However, the cellular components responsible for *oriC* localization are not known. The polar localization of *oriC*-linked sequences remains in *oriC*$^{-}$ strains, which use alternative Hfr replication origins, and it is thought to reflect the recognition of "centromere-like" sequences linked to *oriC* (Gordon *et al.,* 2002). One possibility is that the preference for Ll.LtrB insertion in the chromosomal *ori* and *ter* regions is due to the localization of Ll.LtrB RNPs at the cellular pole, and I investigated this hypothesis in my thesis research. I show that LtrA is in fact pole localized in both *E. coli* and *L. lactis*, and like the *oriC*-linked sequences, the polar localization of LtrA occurs independently of the active replication from *oriC* (Zhao & Lambowitz, 2005). Ll.LtrB retrotransposition sites also remain clustered in *ori* and *ter* region in strains that use alternative origins instead of *oriC* (Beauregard *et al.,* 2006). The polar localization of LtrA could reflect its interaction with host factors that are bound to the origin and/or terminus of DNA replication, with exposed DNA sites in these regions, or with pole-localized cellular components, possibly the same ones that dictate the polar localization of *oriC*-linked sequences.

An increasing number of bacterial proteins have been found to be pole localized (reviewed in Young, 2006). Polar localization of proteins plays a role in mobility and pathogenesis (IcsA, ActA), adaptation to the environment (PleCD, CheA), and cell division and segregation (Min system) (Lybarger & Maddock, 2001; Shapiro *et al.,* 2002).

Despite of the growing list of polar localized proteins, it remains unclear how such polar localization is achieved. An important model system for studying pole localization is the IcsA protein, a polar localized outer-membrane protein mediating actin tail assembly in *Shigella* (Brandon *et al.,* 2003). IcsA, like other proteins secreted via the Sec pathway, contains a well-defined signal peptide (N-terminal 52 aa of IcsA). However, the Sec apparatus is distributed uniformly along the cell membrane and the IcsA protein with signal peptide deleted (IcsAΔSP) still localizes to the pole area in cytoplasm, suggesting that the polar localization of IcsA occurs before secretion and is independent of the signal peptide (Figure 1.7A, Charles *et al.,* 2001; Brandon *et al.,* 2003). Furthermore, two non-overlapping subsegments of IcsA (1-104 aa, 507-620 aa) both displayed polar localization when expressed in both S. *flexneri* and *E. coli*, indicating that the polar information might be redundant in these subsegments. (Figure 1.7B, Charles *et al.,* 2001).

The polar localization of full-length IcsA is altered by mutations in outer membrane lipopolysacchride (LPS) synthesis in *Shigella* (*galU*, *rfe*, Figure 1.7C, Sandlin *et al.,* 1995), or in *E. coli* K-12 strain MBG263 with incomplete LPS (Figure 1.7C, Jain *et al.,* 2006). However, it was suggested by Jain *et al.* that the polar localization of IcsA and other autotranporters (such as SepA of *Shigella*, AIDA-I of diffusely adherent *E. coli*, BrkA of *Bordetella pertussis*) may occur in cytoplasm before translocation to the outer membrane, and the complete LPS is just more rigid in membrane fluidity than incomplete LPS to maintain the polar localization of proteins. Similar to the cytoplasmic polar localization of $IcsA_{507-620}$-GFP fusion in *E. coli* K-12 strain DH10B (Charles *et al.,* 2001), in my research work, I did observe the cytoplasmic polar localization of this fusion in *E. coli* K-12 strain HMS174(DE3), or in LPS synthesis defects *galU* and *rfe* in HMS174(DE3) background (Zhao & Lambowitz, 2005; this work). These results suggest

that the polar localization of proteins in the cytoplasm is independent of the outer membrane LPS composition.

Even in spherical shaped *mreB* mutants, the distribution of IcsA protein is still restricted to several sites in the membrane, which may contain 'polar material or information' (Figure 1.7D, Nilsen *et al.,* 2005). The fact that Min proteins still maintained the MinCDE polar zone in the Δ*mreB* strain, indicates that some 'pole information' for protein localization still exists even in spherical cells (Shih *et al.,* 2005).

Research in this thesis shows that overexpression of Ll.LtrB-encoded protein LtrA abolishes the polar localization of $IcsA_{507-620}$ (Zhao & Lambowitz, 2005), suggesting competition for a common localization determinant in cytoplasm, despite the fact that these two proteins do not have obvious sequence similarities or isoelectric points. It is possible that polar proteins may share the limited 'pole compartment area' or 'pole binding sites', and that different host factors contribute to transporting the proteins to those area or sites.

The identification of such factors would be a major step toward understanding of bacterial protein localization mechanisms, and increase our knowledge to the relationship between protein localization and group II intron mobility, and other fundamental processes in the cell, such as nutrient access, cell division, and motility.

1.6 OVERVIEW OF CHAPTERS

This dissertation focuses on the localization of Ll.LtrB encoded protein – LtrA and its effect on Ll.LtrB intron-insertion preference in *E. coli* genome. Following the introduction in chapter 1, the intracellular localization of LtrA is characterized in detail in chapter 2. The high throughput screening for mutants displaying aberrant LtrA localization and further study of Ll.LtrB intron insertion in these mutants is described in

chapter 3. The last chapter, chapter 4, contains the materials and methods used in this research.

The polar localization of LtrA in *E. coli* provides a possible explanation for Ll.LtrB intron insertion preference in the *ori* and *ter* regions of the *E. coli* chromosome. For localization studies, I constructed vectors expressing GFP-LtrA fusion proteins. In chapter 2, by using the GFP-LtrA fusions as well as immunofluorescence of native LtrA, I show LtrA localizes to cell poles and this polar localization occurs under a broad range of cell growth and induction conditions. I also find similar polar localization patterns of GFP fusions with non-overlapping LtrA subsegments, suggesting that this localization information is redundant or not sequence-specific, possibly involving common physiochemical characteristics of the different polypeptide segments. The GFP/LtrA fusion remains at the poles in *oriC*$^{-}$ strain, indicating that the polar localization of LtrA is independent of active replication.

I discovered that LtrA competes for polar localization of IcsA in *E. coli* and like IcsA, LtrA localizes to the cell poles independently of nucleoid occlusion. The similar localization patterns shared by two proteins with distinct functions makes it of interest to identify the polar localization determinants. In addition, one way to obtain evidence that the preferential insertion of the Ll.LtrB intron into the chromosomal *ori* and *ter* regions is due to the polar localization of LtrA would be to identify mutants with altered LtrA localization patterns and show that they result in correspondingly altered insertional preferences.

The similar localization pattern of LtrA and IcsA provides the rationale to test LtrA localization in the mutants showing aberrant IcsA distribution patterns, as described at the end of chapter 2. LtrA remains at the pole in an *mreB* mutant, suggesting that LtrA, unlike IcsA, does not transport to the pole via MreB scaffold. Additionally, the polar

localization of LtrA is not affected by lipopolysaccharide (LPS) composition in the outer membrane, while the full-length IcsA distribution is more laterally diffused on the outer membrane in both *Shigella* and *E. coli* strains with LPS defects (Sandlin *et al.,* 1995; Jain *et al.,* 2006). In other gene disruptions, such as *hns,* Δ*tat*, *dam*/*dcm*, LtrA localization remains localized at the poles.

In chapter 3, a high throughput screen for host factors affecting LtrA localization identified disruptants in five genes, *gppA, uhpT, wcaK, ynbC*, and *zntR*, in which higher proportion of cells showed diffuse LtrA localization. The distribution of Ll.LtrB intron library insertion sites in all the five mutants is more uniform throughout the genome, demonstrating a correlation between LtrA localization and Ll.LtrB intron insertion site preference. Disruptants *gppA* and *wcaK* also strongly affect wild-type Ll.LtrB intron mobility frequency. Analysis of disruptants *gppA*, and *ppx* and *ppk* in the polyphosphate synthesis pathway, shows diffuse distribution of GFP/LtrA in *ppx* and *gppA* strains, but wild-type polar distribution in *ppk* strain. A high proportion of *gppA* disruptant cells also showed a more diffuse distribution of intracellular polyphosphate similar to the more diffuse distribution pattern of GFP/LtrA. This more diffuse distribution of polyphosphate is also observed in the other four disruptants (*uhpT*, *wcaK*, *ynbC*, and *zntR*) showing diffuse GFP/LtrA localization and the disruptant *ppx*, suggesting that the altered LtrA localization may be related to the polyphosphate accumulation and distribution in the mutant cells.

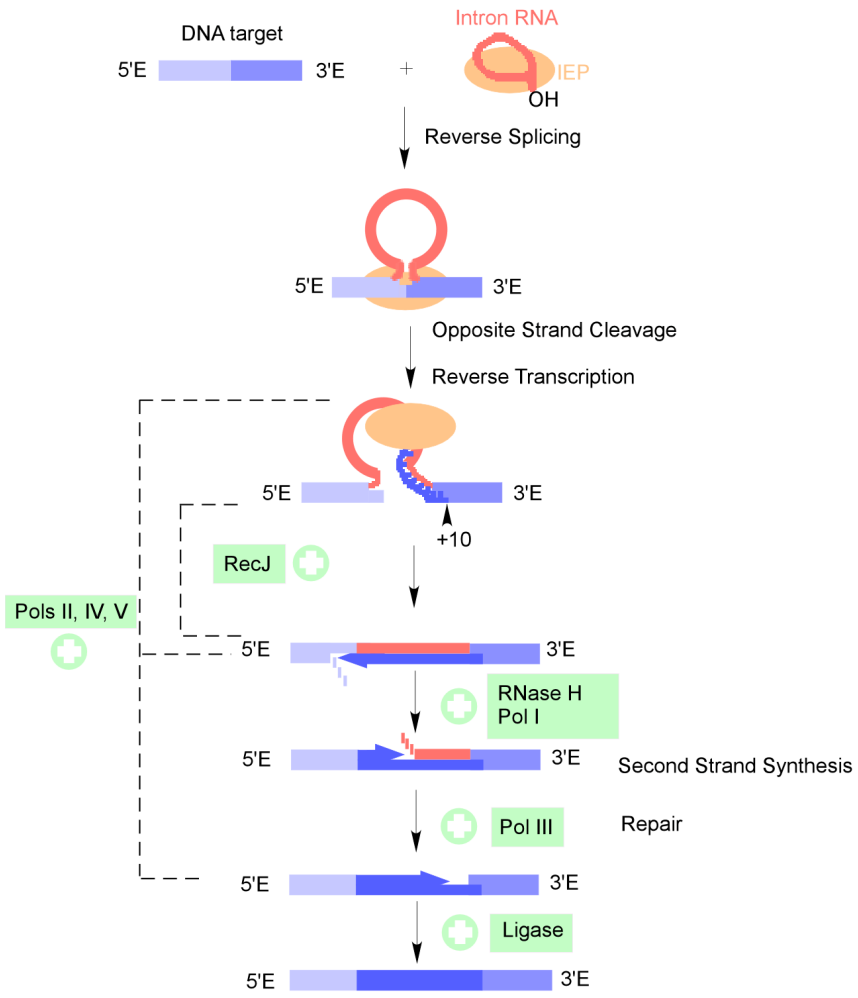

Figure 1.1: Group II intron retrohoming process in *L. lactis* and *E. coli.*

Intron retrohoming (mobility) is accomplished by intron RNA insertion directly into one strand of a DNA target, followed by reverse-transcription of the intron RNA by the intron-encoded protein (IEP). Efficient integration of the intron may require host factors for resection of he bottom DNA strand, permitting extended cDNA synthesis into the 5' exon, intron RNA degradation, second-strand DNA synthesis, and DNA ligation. Intron RNA (red), IEP (peach), DNA strand (different shades of blue), facilitatory host factors (green box, '+'). (Zimmerly *et al.,* 1995a, b; Yang *et al.,* 1996; Eskes *et al.,* 1997; Smith *et al.,* 2005)

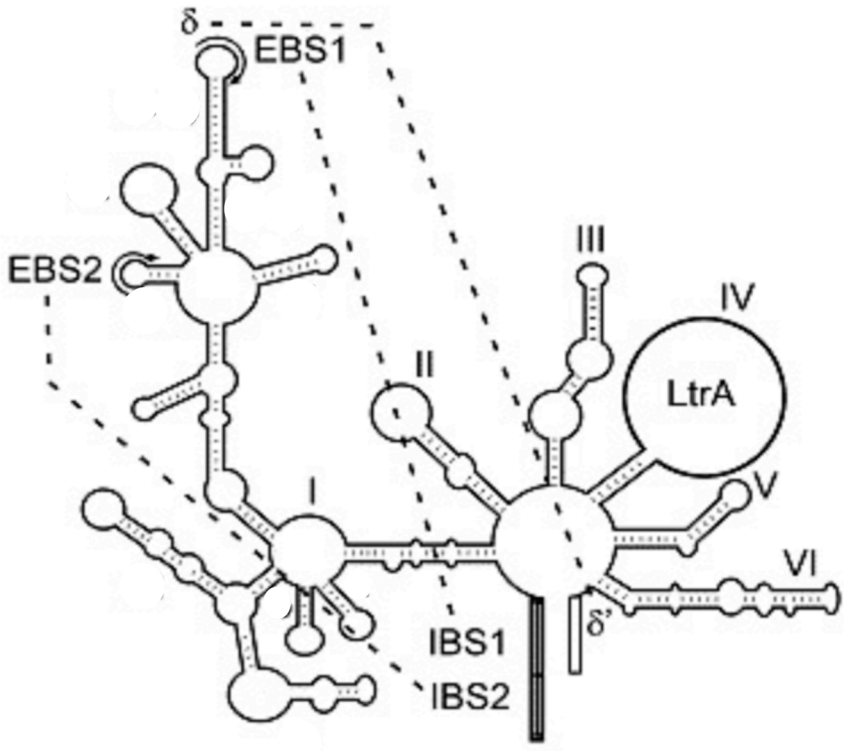

Figure 1.2: Secondary structure and base pairing interactions with flanking exons in the Ll.LtrB intron.

Group II intron Ll.LtrB contains six domains, DI to DVI. Exon-binding sites (EBS1, EBS2, δ) in the intron and intron-binding sites (IBS1, IBS2, δ') in exons are labeled. Base-paring interactions between these sequences are indicated by dashed lines. The LtrA protein is encoded in DIV.

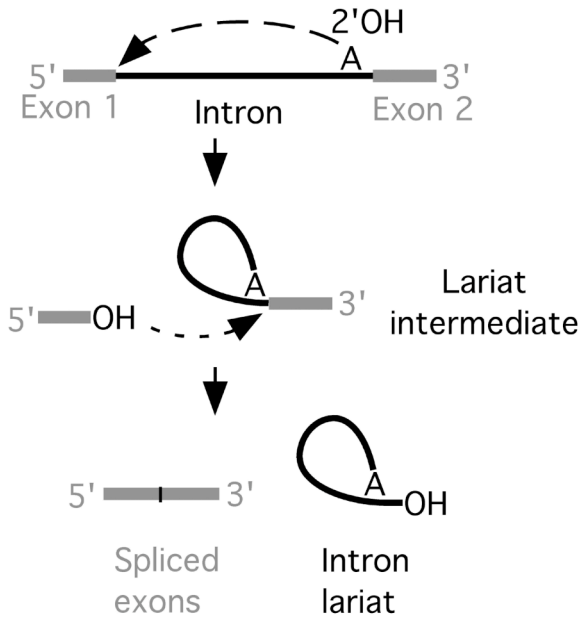

Figure 1.3: Group II intron splicing mechanism.

Splicing is characterized by two sequential transesterification reactions. In the first reaction, nucleophilic attack at the 5' splice site by the 2'-OH group of a bulged A-residue in intron domain VI (DVI), results in cleavage of the 5'-splice site coupled to formation of lariat intermediate. In the second reaction, the 3'-OH of the cleaved 5' exon attacks the 3'-splice site, leading to the exon ligation and the release of the intron lariat. (Michel & Ferat, 1995)

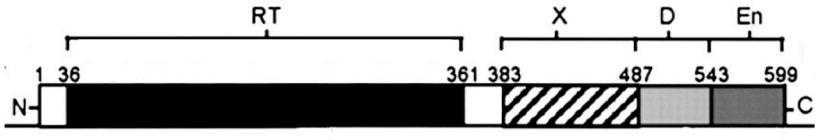

Figure 1.4: The LtrA protein encoded by Ll.LtrB intron.

Four conserved domains in LtrA are shown as rectangular boxes with different shading or patterns: RT, X (maturase, functioning in RNA splicing), D (DNA-binding), and En (endonuclease, essential for bottom strand DNA cleavage in mobility). (San Filippo & Lambowitz, 2002)

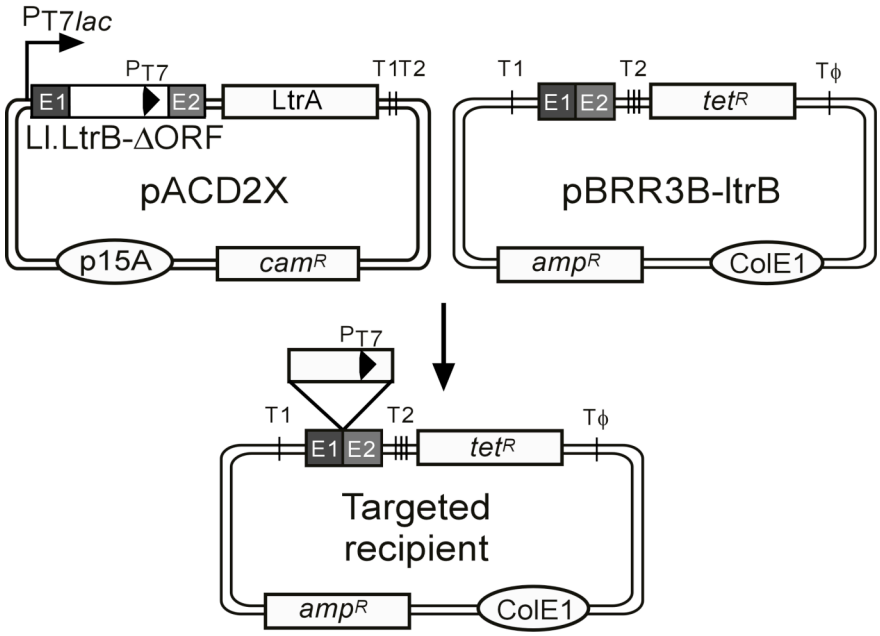

Figure 1.5: Diagram of plasmid-based intron mobility assay.

The donor plasmid pACD2X uses a T7*lac* promoter to express a 0.9-kb Ll.LtrB-ΔORF intron with short flanking exons (E1, E2). The intron contains an additional T7 promoter (P_{T7}) in domain IV. The LtrA ORF is located downstream of the intron. The plasmid is a derivative of the Cam^R vector pACYC184. The recipient plasmid contains the Ll.LtrB target site (ligated E1-E2 sequence, positions -30 to +15 from the intron insertion site) upstream of a promoterless *tet*^{R} gene. The LtrA protein promotes intron RNA splicing and remains bound to the excised intron RNA in RNPs that promote mobility. Insertion of the intron into the target sites brings in the T7 promoter and activates the *tet*^{R} gene. Mobility frequencies are calculated as the ratio of $(Tet^R+Amp^R)/Amp^R$ colonies. The plasmids also contain Tϕ and *E. coli* rrnB T1, T2 terminators as labeled (Guo *et al.*, 2000; Zhong *et al.*, 2003).

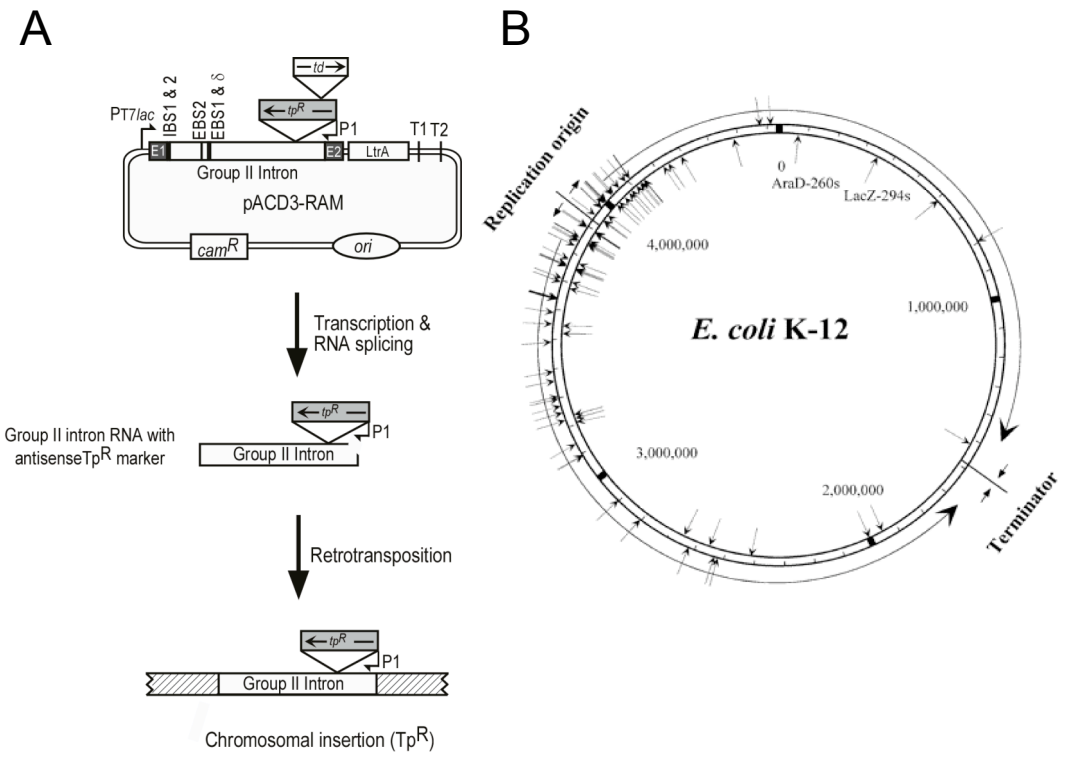

Figure 1.6: Ll.LtrB intron chromosomal targeting and insertion sites distribution in the *E. coli* genome.

(A) Diagram of Ll.LtrB intron chromosomal insertion method. An Ll.LtrB intron library with randomized EBS and d sequences is cloned into pACD3-RAM vector. Tp^R-RAM (retrotransposition-activated selectable marker) in intron domain IV consists of tp^R gene in the opposite orientation to Ll.LtrB intron disrupted by a group I *td* intron in the same orientation as Ll.LtrB intron. During retrotransposition of the Ll.LtrB intron via an RNA intermediate, the group I intron is spliced, reconstituting the tp^R gene, which can be selected by trimethoprim resistance after DNA insertion. (B) Distribution of the Ll.LtrB intron insertion sites in the *E. coli* genome. Insertion sites of the Ll.LtrB intron library with randomized EBS and δ sequences in the genome sequence of *E. coli* K-12 MG1655. 57% of the insertion sites are clustered within 5% of the genome on either side of the bidirectional replication origin (*oriC*) (Zhong *et al.,* 2003).

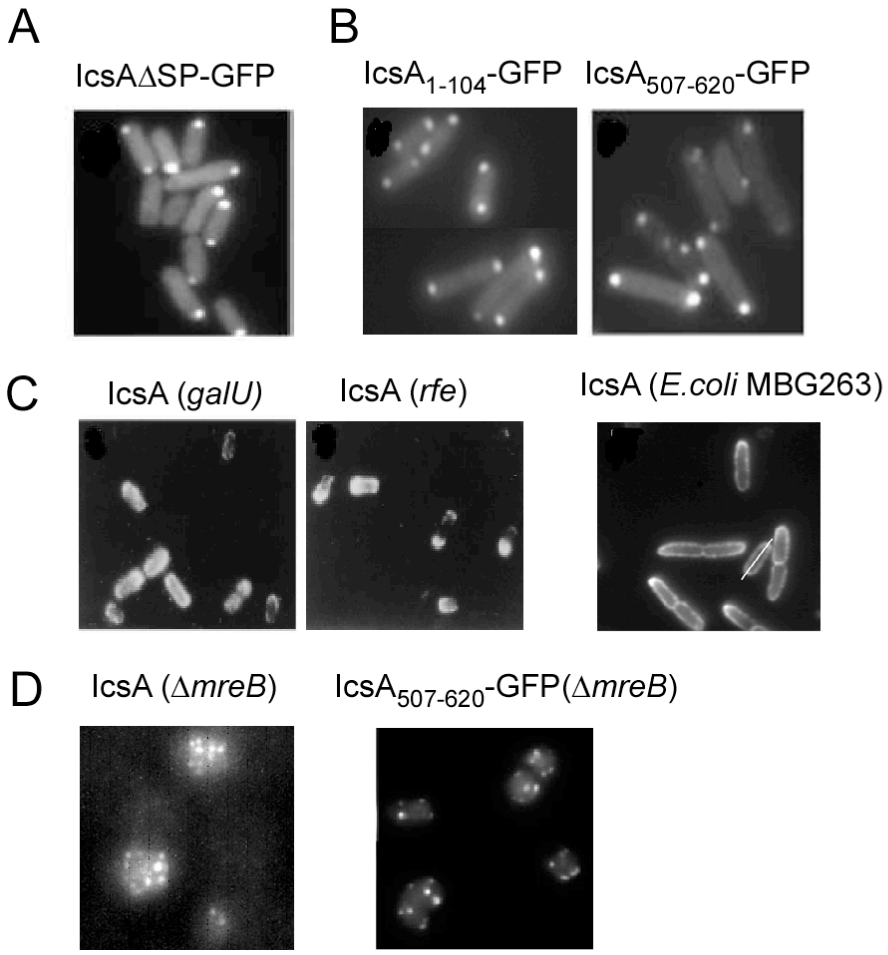

Figure 1.7: Localization patterns of IcsA and IcsA polypeptide segments in *Shigella flexneri* and *E. coli*.

(A) Fluorescence microscopy of C-terminal GFP fusion of IcsA protein without signal peptide (IcsAΔSP) in IcsA$^-$ *Shigella* cells. (B) Fluorescence microscopy of C-terminal GFP fusion to the IcsA segments 1-104 or 507-620 aa in IcsA$^-$ *Shigella* cells. (C) Immunostaining of IcsA in lipopolysaccharide (LPS) synthesis mutants *galU* and *rfe* in *S. flexneri*, or in *E. coli* K-12 strain MBG263 with incomplete LPS. (D) Distribution of full-length secreted IcsA (immunofluorescence) and IcsA$_{507-620}$-GFP fusion in the Δ*mreB E. coli* cells. The data shown are from Sandlin *et al.,* 1995; Charles *et al.,* 2001; Nilsen *et al.,* 2005; Jain *et al.,* 2006.

Chapter 2: Characterization of LtrA localization

The clustering of Ll.LtrB intron insertion sites around the replication origin (*oriC*) in the *E. coli* genome suggests that intron retrohoming events could be influenced by active application, the specific *oriC*-linked sequence, the intercellular localization of host factors in intron mobility, and/or the localization of intron RNPs. Both LtrA and the *oriC*-linked sequence showing similar localization pattern at cell poles in *E. coli*, raises the possibility that LtrA has the higher accessibility to the exposed DNA sites in these genome regions, resulting in the Ll.LtrB insertional preference for the *ori* region. The localization of *oriC*-linked sequence is maintained in *oriC*$^-$ strain, and thus, is independent of active replication (Gordon *et al.,* 2002). Detailed studies of LtrA localization and the distribution of Ll.LtrB intron insertion sites in *oriC*$^-$ strain, would help distinguish the different hypotheses about the etiology of Ll.LtrB insertion site preference, whether the Ll.LtrB intron integration preference is due to the active replication, or the high accessibility of LtrA protein at the same intracellular locus.

2.1 Localization of GFP/LtrA in *E. coli*

2.1.1 Construction of active LtrA-GFP fusions

To study the intracellular localization of the LtrA protein, I constructed N- and C-terminal LtrA-GFP fusions in the intron-donor plasmid pACD2X (pACD2X-GFP/LtrA, pACD2X-LtrA/GFP, Figure 2.1A). This plasmid contains a 0.9-kb Ll.LtrB-ΔORF intron with short flanking exons cloned behind a T7*lac* promoter. The LtrA ORF or LtrA fused to GFP is expressed from a position just downstream of the 3' exon. The intron has an additional T7 promoter inserted near its 3' end for use in intron mobility assays as

described in detail in chapter 1. Briefly, the LtrA protein expressed from the downstream *cis* position promotes RNA splicing and then remains tightly bound to the excised intron RNA in RNPs that promote intron mobility. In the intact Ll.LtrB intron and in pACD2X, the synthesis of LtrA is autoregulated by binding to its own Shine-Dalgarno sequence, thereby limiting the accumulation of excess unbound protein (Singh, *et al.,* 2002). SDS-PAGE and immunoblotting with an anti-LtrA antibody showed that both the N- and C-terminal LtrA-GFP fusions were expressed at somewhat reduced levels (33-50% of wild type), but the proportion of expressed protein recovered in RNPs was essentially the same as for wild-type LtrA (data not shown).

To determine whether the LtrA-GFP fusions are active, I carried out intron mobility assays in which Cam^R donor plasmids expressing wild-type LtrA or the LtrA-GFP fusions were cotransformed into *E. coli* HMS174(DE3) with a compatible Amp^R recipient plasmid containing the Ll.LtrB target site cloned upstream of a promoterless tet^R gene (Figure 2.1A). After induction of donor plasmid expression with IPTG (isopropyl-β-D-thiogalactopyranoside), insertion of the intron carrying the T7 promoter into the target sites activates the expression of the tet^R gene, and mobility frequencies are measured as the ratio of $(Tet^R+Amp^R)/Amp^R$ colonies. Both the N- and C-terminal LtrA-GFP fusions supported mobility at frequencies that were roughly proportional to their somewhat lower expression levels (Figure 2.1B, mobility frequencies 22 ± 1% and 23 ± 3% for pACD2X-GFP/LtrA and pACD2X-LtrA/GFP, respectively, compared to 85 ± 3% for wild-type LtrA).

2.1.2 Intracellular localization of LtrA-GFP fusions

Next I examined the localization of the LtrA-GFP fusions by fluorescence microscopy in *E. coli* HMS174(DE3) grown and induced under similar conditions as in the mobility assay (250 μM IPTG for 1 h at 37°C, Figure 2.2A, B, Table 2.1). Most of the

cells expressing either the GFP/LtrA or LtrA/GFP fusions showed a localization pattern with two foci at the poles or two foci at the poles plus an additional focus toward the middle of the cell (82.3 and 53% for pACD2X-GFP/LtrA and pACD2X-LtrA/GFP, respectively). Cells with two foci were smaller (3.1 ± 0.47 μm) than those with three foci (5.2 ± 0.57 μm), suggesting that the appearance of the third focus is correlated with incipient cell division. Lower proportions of the cells showed only one focus at a pole (1.3 and 2.0% for pACD2X-GFP/LtrA and pACD2X-LtrA/GFP, respectively), or four or more foci, with two at the poles and the remainder distributed throughout the cell (16.4 and 35.0% for pACD2X-GFP/LtrA and pACD2X-LtrA/GFP, respectively). Controls showed that GFP expressed under the same conditions from a parallel construct was uniformly distributed throughout the cell, as expected (pAC-GFP, Figure 2.2C).

In other experiments, I also observed pole localization of the N-terminal LtrA-GFP fusion expressed from the normal location within the intron (pACDF-GFP/LtrA, Figure 2.2D, Table 2.1) and for the "protein only" construct, which lacks the Ll.LtrB intron (pAC-GFP/LtrA, Figure 2.2E, Table 2.1). The GFP/LtrA fusion was also pole localized in *E. coli* strains BL21(DE3) and DH5α (not shown). Additionally, I found that pole-localized GFP/LtrA expressed at lower levels from the *lac* promoter (plac-GFP/LtrA, Figure 2.2F, Table 2.1) could be competed from the poles by the full-length LtrA protein expressed from pACD2X (Figure 2.2G), but not by maltose-binding protein expressed at similar levels and confined intracellularly by deletion of its signal peptide (pAC-MBP, Figure 2.2H).

Staining of DNA with DAPI and cell membranes with FM 4-64 showed that pole-localized GFP/LtrA is located at the edge of, but mostly excluded from the nucleoid region, and generally apposed to the inside of the inner membrane (shown in Figure 2.2E for pAC-GFP/LtrA).

2.2 LOCALIZATION OF GFP/LTRA IN *L. LACTIS*

To investigate the intracellular localization of LtrA in *L. lactis*, I modified an existing construct, denoted pLE-RIG (Ichiyanagi *et al.,* 2002), to express an N-terminal LtrA-GFP fusion. The GFP protein I used here was a variant of GFPuv, named GFPrft (obtained from Dr. George Georgiou, Univ. of Texas Austin), which produces detectable fluorescence in a low pH environment, as in case for *L. lactis* (Fernandez de Palencia *et al.,* 2000). Compared to GFPuv, a higher proportion of the acidic *Lactococcus* cells showed fluorescent GFPrft (data not shown). The modified construct (pLE-RIG-GFP/LtrA; Figure 2.3C), contains the full-length Ll.LtrB intron and short flanking exons cloned behind an inducible *nisA* promoter, with the GFP/LtrA fusion protein expressed from the native location within intron domain IV.

After transformation into *L. lactis* strain NZ9800 and induction with nisin, 90% of the cells showed one (35%) or two foci (55%) apposed to or near the membrane on opposite sides of the cell, with an additional 5% showing two foci on opposite sides plus a third focus in the middle; the remaining 5% of cells showed one focus elsewhere in the cell (Figure 2.3A, Table 2.1). Although many of the cells were roughly spherical, pole localization could be clearly discerned in those cells that were somewhat elongated or growing in chains. Thus, the intracellular localization of the GFP-LtrA fusion in *L. lactis* appears analogous to that in *E. coli*. By contrast, GFP expressed by itself gave uniform fluorescence throughout the cell (Figure 2.3B).

2.3 IMMUNOFLUORESCENCE OF LTRA

2.3.1 Immunofluorescence of LtrA in *E. coli*

To exclude the possibility that the pole localization of the GFP/LtrA fusion protein is an artifact resulting from the disruption of normal localization signals by the

GFP fusions, I examined the localization of LtrA by immunofluorescence microscopy. These experiments used HMS174(DE3) and the RNP expression construct pACD2X (Figure 1.5 in Chapter 1). After IPTG-induction, the cells were fixed and probed with an anti-LtrA antibody preparation, followed by IgG-FITC secondary antibody. As shown in Figure 2.4A, the LtrA protein detected in this assay had essentially the same polar localization pattern seen for GFP/LtrA fusions, with two foci at the poles in smaller cells and an extra focus in the middle in larger cells. By contrast, untransformed control cells showed only low background fluorescence (data not shown).

2.3.2 Immunofluorescence of endogenous LtrA in *L. lactis*

I also attempted to detect LtrA synthesized from the endogenous chromosomal copy of Ll.LtrB in *L. lactis* strain NZ9800 strain by immunofluorescence microscopy with anti-LtrA antibody. Immunoblots showed that the level of LtrA expressed from the endogenous element was <2% that from nisin-induced pLE-RIG-GFP/LtrA (data not shown). The immunofluorescence microscopy showed correspondingly very light foci with the same localization pattern found for the GFP/LtrA fusion (Figure 2.4B). These foci were quickly bleached, but not observed when the primary anti-LtrA antibody was omitted, nor in NZ9800*ΔltrB*, which is deleted for the endogenous Ll.LtrB intron (Ichiyanagi *et al.,* 2002; data not shown). Thus, despite qualifications required by the low expression level, LtrA synthesized from the endogenous integrated element also appears to be pole localized.

2.4 CHARACTERIZATION OF GFP/LTRA LOCALIZATION IN *E. COLI*

2.4.1 Polar localization occurs at different levels of LtrA expression

I next tested whether the localization pattern of GFP/LtrA fusions might be affected by their expression level. For these experiments, I used both the RNP expression

construct pACD2X-GFP/LtrA and the protein-only expression construct pAC-GFP/LtrA at different temperatures (25, 30 and 37°C), IPTG concentrations (50 or 250 μM), and induction times (60 to 180 min). Similar results were obtained for both constructs (Table 2.1, Figure 2.5; data not shown). Under all conditions tested, polar localization was seen in >95% of the cells. Further, the same polar localization was seen when LtrA was expressed at a lower level from the *lac* promoter (plac-GFP/LtrA) with IPTG induction or by "leaky" expression without IPTG induction (Figure 2.5). The pole localization pattern seen for LtrA is not expected for inclusion bodies which are generally distributed throughout the cell (Carro *et al.,* 2005), and the phase contrast microscopy images of cells expressing GFP/LtrA do not show inclusion bodies and are indistinguishable from those of cells without expression vector (not shown).

2.4.2 Polar localization of LtrA occurs over a wide range of growth rates

Coros *et al.* (2005) studying the retrotransposition of Ll.LtrB to ectopic sites in *E. coli* found that the clustering of insertion sites in the *ori* and *ter* regions was most pronounced (93% of insertions) in slowly growing cells induced with high IPTG concentrations (1 mM; doubling time 60 min at 37°C), and somewhat less pronounced (80% of insertions) in more rapidly growing cells induced with lower IPTG concentrations (100 μM; doubling time 23 min at 37°C). I found that LtrA remained predominantly pole localized at IPTG concentrations ranging from 50 μM to 1 mM (60 min induction at 37°C), where doubling times ranged from 30 to 85 min (Table 2.1).

2.4.3 Localization of GFP-fusions with different LtrA subsegments

To determine whether a specific region of LtrA is responsible for its polar localization, I tested GFP fusions with different subsegments of the protein. All LtrA subsegments tested, including three non-overlapping regions (amino acid residues 2-200,

201-400, and En), showed pole localization patterns similar to that of full-length LtrA (Figure 2.6, Table 2.1). In other experiments, I also found polar localization from the *Neurospora crassa* mitochondrial tyrosyl-tRNA synthetase (CYT-18 protein) (Kittle *et al.,* 1991), which is unrelated to LtrA but has similar size (637 amino acids) and basicity (PI = 9.29, LtrA PI =9.64). However, CYT-18 is closely related to *E. coli* TyrRS, whose intracellular localization is unknown. Together, these results suggest either that signals responsible for the polar localization are common and redundant or that some physical property of the protein (*e.g.*, positively charged regions) is responsible for the polar localization.

2.5 LOCALIZATION OF LTRA IN AN *ORIC*$^-$ STRAIN

The polar localization of LtrA could account for the clustering of Ll.LtrB-insertion sites in the *ori* and *ter* regions of the *E. coli* chromosome, which are similarly localized for much of the cell cycle (Niki & Hiraga, 1998; Niki *et al.,* 2000; Draper & Gober, 2002). The *oriC*-linked sequences remain polar localized in *oriC*$^-$ strains, independent of active replication from *oriC* (Gordon *et al.,* 2002). To test whether or not LtrA's polar localization is dependent upon the functioning of *oriC*, I examined the localization of the GFP/LtrA fusion in the *oriC*$^-$ strain AQ10060(DE3) and its *oriC*$^+$ parent AQ10033(DE3) (Figure 2.7). Because the *oriC*$^-$ strain AQ10066 is Amp^RCam^R (minimal *oriC* is replaced an amp^R gene plus a cam^R gene in the *asnA* gene next to the amp^R gene), I constructed an alternative intron expression plasmid pACSD2-GFP/LtrA carrying a spc^R marker.

In the *oriC*$^+$ strain AQ10033(DE3), the GFP/LtrA fusion showed the same polar localization pattern as in HMS174(DE3), with the majority of cells having two foci near the poles (Figure 2.7A, B). Most (72.1%) of the *oriC*$^-$ AQ10060(DE3) cells were of normal size, and in these cells, GFP/LtrA showed polar localization (Figure 2.7C, Table

2.1), which, like the polar localization of *oriC*-linked sequences, is not dependent upon the function of *oriC*. Additionally, *oriC*$^-$ cells have an increased tendency to form filaments in which *oriC*-linked sequences are found at multiple foci throughout the cell (Gordon *et al.,* 2002). In our experiments, 27.9% of the *oriC*$^-$ cells formed filaments, and these likewise had multiple GFP/LtrA fluorescence foci distributed throughout the cell (two examples are shown in Figure 2.7D). Thus, the disrupted processes that lead to filament formation and mislocalization of *oriC*-linked sequences in some *oriC*$^-$ cells may also lead to mislocalization of LtrA.

Similar to the distribution preference of Ll.LtrB intron retrohoming insertion sites in the *ori* region of the *E. coli* chromosome (Zhong *et al.,* 2003), the distribution of intron retrotransposition sites is clustered in both *ori* and *ter* region (Coros *et al.,* 2005). Beauregard *et al*. (2006) observed that this clustered distribution of retrotransposition sites also remained in an *E. coli* Δ*oriC* host. Together, both LtrA localization and intron insertion site preference are independent of active replication, suggesting that the polar localization of LtrA might provide higher accessibility of Ll.LtrB intron to the DNA sites in sequences localizing in the same intracellular area and increase the intron insertion efficiency in those regions of the chromosome.

2.6 LTRA INTERFERES WITH POLAR LOCALIZATION OF THE *SHIGELLA* ICSA PROTEIN

2.6.1 Polar localization of IcsA and IcsA-GFP fusion

The *Shigella* outer membrane protein IcsA (VirG) localizes to the old pole of the bacterium to mediate polarized assembly of an actin tail that pushes the bacterium through the cytoplasm of infected mammalian cells and into adjacent cells (Goldberg *et al.,* 1993; Janakiraman & Goldberg, 2004). In *E. coli*, GFP fusions of two non-overlapping IcsA fragments ($IcsA_{1-104}$ and $IcsA_{507-620}$) both showed polar localization

patterns (Charles *et al.,* 2001). Among proteins reported in the literature to be pole localized in *E. coli*, the localization patterns of $IcsA_{1-104}$ and $IcsA_{507-620}$ appeared to be particularly similar to that of LtrA. To investigate a possible link between the polar localization of LtrA and that of IcsA, I obtained the construct pBAD24-*icsA*$_{507-620}$*::gfp* from Dr. Marcia Goldberg (Massachusetts General Hospital, Cambridge, MA). In addition, using IcsA amplified from a chromosomal DNA template provided by Dr. Shelley Payne (University of Texas at Austin), I constructed plasmids pACD2X-IcsAΔSP to express full-length IcsA without its signal peptide.

2.6.2 LtrA interferes the polar localization of IcsA in *E. coli*

To test whether $IcsA_{507-620}$ and LtrA compete for pole localization determinants, the two proteins were co-expressed in *E. coli* HMS174(DE3). The $IcsA_{507-620}$-GFP fusion protein expressed by itself showed the expected polar localization pattern, but with the proportion of cells containing only a single focus at one pole higher than that for LtrA (Figure 2.8A, Table 2.1). Significantly, the coexpression of LtrA interfered with the pole localization of the $IcsA_{507-620}$-GFP fusion, resulting in a high proportion of cells (70.5%) showing diffused fluorescence (Figure 2.8A, B, Table 2.1). The remaining cells showed pole localization (22.8%) or one fluorescent focus elsewhere (6.7%). I note that about a third of cells tabulated as showing pole localization in this experiment also had high dispersed background fluorescence, suggesting incomplete interference. Immunoblots with anti-GFP antibody showed that the coexpression of LtrA reduced $IcsA_{507-620}$-GFP expression by only about 1/3 (not shown). In reciprocal experiments, full-length IcsA with its signal peptide deleted (IcsAΔSP, Figure 2.8C, D) could not displace LtrA from the poles. These findings suggest that LtrA may have a higher affinity for a required positional determinant than IcsA.

2.6.3 Localization of LtrA or IcsA is independent of nucleoid occlusion

Janakiraman and Goldberg (2004) obtained further insight into the mechanism of IcsA localization by treating *E. coli* with aztreonam to inhibit the cell division protein FtsI. In the resulting filamentous cells, $IcsA_{507-620}$ foci were no longer confined to the poles, but also appeared at regularly spaced intervals both between nucleoids and in anucleate segments, with the spacing between foci suggesting localization to potential cell division sites. I found that GFP-LtrA foci behaved similarly in filamentous aztreonam-treated HMS174(DE3), with a spacing between foci of 3.4 ± 1.0 μm compared to 3.0 ± 0.8 μm for pole localized foci in untreated cells (Figure 2.9A). Notably, as for IcsA, this spacing was maintained in anucleate segments that appeared in ~5% of the filamentous cells (Figure 2.9A, right panels). Additionally, the polar localization of GFP/LtrA remained in untreated HMS174(DE3) after intracellular DNA was largely digested by DNase I in the presence of lysozyme (Figure 2.9B). These findings indicate that the localization of GFP/LtrA is not the result of nucleoid occlusion and are consistent with localization to potential cell division sites, although the factors responsible for such polar localization of proteins without homology in sequence or function remain unidentified.

2.7 LtrA localization is maintained in host factor mutants affecting cell rod shape, membrane fluidity, nucleoid condensation, and the twin-arginine protein translocation pathway (TAT)

Any host factors involved in a polar localization mechanism might display the mislocalized LtrA protein in the corresponding mutants or knockouts. After characterizing LtrA localization in *E. coli* and *L. lactis*, I examined the LtrA distribution in a number of *E. coli* mutants for host factors that might affect polar localization to further understand the polar localization mechanism of LtrA and perhaps other proteins

with similar distribution pattern. Since LtrA interferes with the polar localization of IcsA segments and may share the same 'pole binding sites' in *E. coli*, I started my examination of LtrA localization in mutants known to show aberrant distribution pattern of IcsA.

2.7.1 LtrA localization in spherical *mreB* mutant

The actin homologue MreB functions in chromosome segregation, cell rod-shape maintenance, and cell polarity in *E. coli* cells (Kruse *et al.,* 2005). Nilsen *et al.* (2005) found that IcsA-GFP fusion proteins (full length or $IcsA_{507-620}$) showed multiple foci in the spherical cells of an *mreB* disruptant identified by screening transposon-induced mutants (Nilsen *et al.,* 2005). This finding suggested that cells that lose their rod shape may lose some determinants of polar localization. MreB, however, is not the only factor that determines cell polarity and protein polar localization. Min proteins still maintained the MinCDE polar zone and the axis of oscillation of MinD in Δ*mreB* cells (Shih *et al.,* 2005), indicating that some determinants of polar localization exists even in spherical cells.

I examined the GFP/LtrA localization in an *mreB* mutant (WM2001, obtained from Dr. Marlene Belfort, Wadsworth Center, Albany, NY) and found that, unlike IcsA which formed many foci, GFP/LtrA typically formed one or two foci in the spherical *mreB* mutant cells (Figure 2.10). Although a few large cells (~2%) did show more than two GFP/LtrA foci, these could be categorized more as filaments rather than normal cells. Excluding these cells, the GFP/LtrA localization pattern in the *mreB* mutant is similar to that in the wild-type strain, with one (34.6%) or two foci (26.5%) localized at opposite sides of the cell (Figure 2.10, Table 2.2).

Confocal microscopy was used to examine the GFP/LtrA localization in the *mreB* mutant and to investigate the organization of foci beneath the cell surface and position relative to the cell membrane. The cell membrane was stained with FM4-64 to show the

relative location of GFP/LtrA (green) to membrane (red, Figure 2.10C). Shown in Figure 2.10C is a representative *mreB* mutant cell (WM2001) with two foci adjacent to but not attached to the cell membrane. Similar to the maintained localization pattern of GFP/LtrA at opposite sides of cells, intron retrotransposition sites in this *mreB* mutant strain remained clustered in the *ori* and *ter* domains as in wild type (Beauregard *et al.,* 2006). These data indicated that LtrA localization is not dependent upon the cell is rod-shape. Although LtrA interferes with the polar localization IcsA, possibly sharing the same polar receptor sites, LtrA and IcsA are affected somewhat differently in this *mreB* mutant.

2.7.2 Localization of LtrA in mutants with LPS synthesis defects

In *Shigella flexneri,* mutations in the *galU* (glucose-1-phosphate uridylyltransferase) and *rfe* (undecaprenyl-phosphate alpha-N-acetylglucosaminyltransferase) genes, which are involved the lipopolysacchride (LPS) biosynthesis, disrupt the polar localization of full-length IcsA protein, leading to a more laterally diffused distribution (Sandlin *et al.,* 1995). Jain *et al.* (2006) recently reported that full-length IcsA also diffuses around the outer membrane in *E. coli* K-12 strain MBG263 with incomplete LPS, but secretion incompetent IcsA with a GFP fusion (from construct pBAD24-$icsA_{1\text{-}24/53\text{-}757}$::*gfp*) remains pole localized in the bacterial cytoplasm. The above results suggest that the polar localization of IcsA protein in the cytoplasm is independent of LPS, but that complete LPS helps to maintain the polar localization in the outer member after secretion.

To test whether LtrA localization is related to LPS in *E. coli*, we used targetrons to disrupt the *galU* (targetron Gal-256a) and *rfe* (targetron Rfe-252s) genes in *E. coli* HMS174(DE3). In both disruptants, GFP/LtrA showed the same cytoplasmic polar localization pattern as in wild-type HMS174(DE3) cells (72.5% and 69.8% of cells showing one or two foci at poles in the *galU* and *rfe* disruptants respectively, Figure 2.11,

Table 2.2), under the induction condition at 500 μM IPTG at 30ºC overnight. The distribution of $IcsA_{507-620}$-GFP fusion was also examined in these mutants. Unlike the diffused localization pattern of full-length IcsA in the corresponding *Shigella* mutants, I found that cytoplasmically localizated $IcsA_{507-620}$ remained at the poles in the *E. coli galU* and *rfe* disruptants and in wild-type K-12 strain HMS174(DE3) (Figure 2.11). Localization of GFP/LtrA and $IcsA_{507-620}$-GFP in the cytoplasm, but not the outer membrane in these *E. coli* cells, would explain the lack of effect of the outer membrane fluidity change in both LPS mutants, although the effect of inner membrane composition on LtrA remains unclear.

In other experiments (Figure 2.12, Table 2.2), I also showed that the polar localization pattern of LtrA in *E. coli* is not dependent upon factors involving chromosome nucleoid condensation (*hns*, *stpA*, *hns*/*stpA*, obtained from Dr. Marlene Belfort), methylation (*dam*/*dcm* strain ER2925 requested from New England Biolabs), DNA replication (IHF subunit *himA* disruptant, generated by designed Ll.LtrB targetron HimA-140a), ATP-dependent helicase SrmB and HelD (generated Wendy Wang by targetrons in our lab), the type IV twin-arginine translocation (Tat) pathway (Δ*tatC*, obtained from Dr. George Georgiou), and cell membrane phospholipids composition (*pgsA*$^{-}$, *pssA*, obtained from Dr. William Dowhan, University of Texas-Houston Medical School). Similar results were also observed in *hns*, *stpA*, *mukB, mreB,* and *seqA* mutant strains by Beauregard *et al*. (2006). In conclusion, the inability to identify host factors by examining the LtrA localization in individual mutants demands a more efficient high throughput screening method to identify host mutants affecting LtrA localization (see chapter 3).

2.8 DISCUSSION

2.8.1 Characterization of LtrA localization

I found that a group II intron-encoded RT, the LtrA protein encoded by the Ll.LtrB intron, is localized to cellular poles in both *E. coli* and *L. lactis*. The pole localization in *E. coli* is observed over a wide range of cellular growth rates and LtrA expression levels, and occurs with or without co-expression of the Ll.LtrB intron RNA, which assembles with LtrA into RNPs that mediate intron mobility. The ability of LtrA to compete with IcsA for pole localization suggests that LtrA localization is dictated by a physiologically relevant mechanism involving interaction with a limited number of localized cellular binding sites that can be occupied either by LtrA or IcsA. Our results and those for IcsA are compatible with the idea that these binding sites are protein or other type of receptor molecules located in the cytoplasm or on the cytoplasmic face of the inner membrane (Janakiraman & Goldberg, 2004). It may be pertinent that in mitochondrial systems, there have been persistent speculative indications that group II intron splicing factors are associated with the inner membrane (Slater *et al.,* 1995).

Two non-overlapping segments of IcsA pole-localize independently, suggesting that they may contain redundant localization signals (Charles *et al.,* 2001). In my research, I observed pole localization for three non-overlapping segments of LtrA, and for the unrelated *N. crassa* mitochondrial tyrosyl tRNA synthetase (CYT-18 protein). These findings could reflect that pole localization of LtrA as well as IcsA is dictated by common, redundant signals or some physiochemical property of the proteins, such as regions of high basicity. The amino acid sequences responsible for pole localization of IcsA are not known. Computer analysis (http://meme.nbcr.net/meme/website/intro.html) revealed a number of short sequence motifs of varying stringencies that are common to

$IcsA_{507-620}$ and the three localized LtrA subsegments, but their significance, if any, remains to be evaluated.

I also found that the pole localization of LtrA is independent of *oriC* function, suggesting that it is not dictated by interaction with DNA structural features or protein components associated with active replication origins (*e.g.*, hemi-methylated DNA or SeqA; Slater *et al.,* 1995). It is possible that the pole localization of LtrA and perhaps other proteins found at the cellular poles is dictated by interaction with the same cellular machinery responsible for the pole localization of the *ori* region. This possibility is consistent with the appearance of a third LtrA focus at midcell in larger cells, which may be about to undergo cell division, as well as the altered localization of both LtrA and *oriC*-linked sequences in filamentous cells derived from $oriC^-$ mutants (Figure 2.7D, Gordon *et al.,* 2002). It is also possible that LtrA localization is dictated by interaction with membrane components that pole localize independently of the *ori* region.

2.8.2 Polar localization of LtrA in mutants with aberrant IcsA localization

I observed the polar localization of LtrA in *E. coli,* adding one more protein to the increasing number of proteins occupying polar sites (Shapiro, *et al.,* 2002; Janakiraman & Goldberg, 2004). Polar localization has been found for proteins with diverse functions in flagella, chemotaxis, type II, III, IV secretion, auto-transportation, and chromosomal segregation, suggesting the importance of polarity in many different aspects of the cells's life cycle. It also indicates that the mechanisms dictating polar localization may be very complex. Among all the reported polar proteins, IcsA showed a particularly similar localization pattern to LtrA. Coexpression of LtrA abolished the polar localization of IcsA, but not *vice-versa*. This result suggests that LtrA and IcsA might compete for the same polar binding sites in the cells, and LtrA might have higher affinity to these sites than IcsA.

It was reported that in *Shigella*, two mutants in the lipopolysaccharide (LPS) synthesis pathway (*galU* and *rfe*) showed a more laterally diffused localization of IcsA (Sandlin *et al.,* 1995). Failure to observe this pattern for $IcsA_{507-620}$-GFP or GFP/LtrA in the same *E. coli* mutants suggested that LtrA is localized within the cytoplasm and not affected by outer membrane fluidity.

Unlike rod-shaped cells, spherical cells may seem to be symmetrical and do not have any polarity. Some polar localized proteins, such as IcsA and EpsM, do not show the typical localization patterns with one or two foci in the spherical cells (Nilsen *et al.,* 2005). However, loss of polar localization does not result in a uniformly distributed localization of these proteins. The Ll.LtrB encoded protein, LtrA, which is polar localized in rod-shaped *E. coli* cells, also localizes as one or two foci in *L. lactis* along the elongation axis of the chain (Zhao & Lambowitz, 2005). In the spherical *mreB* mutant of *E. coli*, the majority of the cells still showed one or two foci at the opposite edge of the cells (this work and Beauregard *et al.,* 2006). This suggests that although there is no 'pole' in the spherical cells, some unidentified information dictating polarity still exists at some with the cell.

2.8.3 Effect of LtrA localization on Ll.LtrB intron target site distribution

Although the polar localization of LtrA can account for the preferential insertion of Ll.LtrB introns with randomized EBS and δ sequences in the *ori* region of the *E. coli* chromosome, it is clearly not the only factor that contributes to dictating integration sites. Coros *et al.* (2005) found that the very strong clustering of Ll.LtrB retrotransposition sites in the *ori* and *ter* regions was modulated somewhat in more rapidly growing cells, while I find LtrA remains largely pole localized in both slowly and rapidly growing cells. Further, Coros *et al.* (2005) characterized retrotransposition events as occurring by DNA-endonuclease dependent or -independent pathways, depending on whether or not the

target site could support second-strand cleavage by LtrA's En domain. The predicted En-dependent events showed a bias for both the *ori* and *ter* regions, while the predicted En-independent events, which may require nascent strands at DNA replication forks to prime reverse transcription, favored the *ori* region, with a gradient toward the *ter* region. Additionally, Ll.LtrB retrotransposition sites were found to be uniformly distributed throughout the chromosome in *L. lactis* (Ichiyanagi *et al.,* 2002), where I found LtrA is localized to discrete foci at opposite ends of the cell. These findings suggest that chromosome packaging and access to DNA replication forks also play a role in dictating group II intron insertion sites and that the relative contribution of these factors may differ significantly between *E. coli* and *L. lactis* (see also Coros *et al.,* 2005).

Ll.LtrB RNPs bind DNA nonspecifically and then, search for target sites by facilitated diffusion along DNA similar to mechanisms used by site-specific DNA-binding proteins (Aizawa *et al.,* 2003). The pole localization of LtrA in *E. coli* may concentrate group II intron RNPs in proximity to exposed DNA segments in the *ori* and *ter* regions, thereby facilitating their initial non-specific DNA binding, after which the RNPs can search for target sites along the DNA. This scenario readily accounts for the preferential insertion of Ll.LtrB insertions sites in the *ori* and *ter* regions, while the distal regions may be more or less accessible to the RNPs depending on growth conditions. The polar regions may also be favorable sites for interaction with components of the DNA replication machinery, such as Pol III, which is pole localized in non-replicating cells (Onogi *et al.,* 2002). Such interactions may facilitate access of group II intron RNPs to DNA replication forks and/or minimize the interval between the initial steps of intron mobility and downstream steps, such as second-strand synthesis, which may depend upon host DNA replication (Smith *et al.,* 2005). Finally, the pole localization of LtrA could potentially link group II intron mobility to the cell division machinery and/or facilitate

the segregation of group II intron RNPs to daughter cells. The latter could be particularly beneficial for many group II introns found in plasmids, which may themselves have unreliable segregation mechanisms (Ichiyanagi *et al.,* 2003).

2.9 SUMMARY

In *E. coli*, Ll.LtrB group II intron shows a pronounced preference for insertion into the *ori* and *ter* regions of the genome. By using fluorescence microscopy, I found that LtrA is localized at the cellular poles, where the *ori* and *ter* regions are located during much of the cell cycle. These results raise a possibility that the interaction between LtrA and components that are bound to the origin of DNA replication might exposes DNA sites in these regions, facilitates chromosomal targeting of the Ll.LtrB intron and results in the *ori* and *ter* regions preference. This polar localization pattern of LtrA exists under various growth or induction conditions. Three non-overlapping subsegments of LtrA showed the similar localization pattern at poles, suggesting that the localization information might be redundant. Both GFP/LtrA and *oriC*-linked sequence remains localized at the poles in an *oriC*$^-$ strain suggesting that this localization is independent of active DNA replication from origins. The ability of LtrA to interfere with the polar localization of an IcsA fragment, but not vice versa, indicated that LtrA may share the similar localization mechanism and/or share the limited 'pole compartment area' or 'pole binding sites', and LtrA may have higher affinity to these 'pole sites' than IcsA. Studies of GFP/LtrA localization in various *E. coli* mutants showed that the polar localization of LtrA is independent of cell shape, membrane fluidity, nucleoid condensation, and twin-arginine protein translocation apparatus.

Table 2.1: Localization of GFP fusion proteins.

Strains and constructs	°C/μM IPTG/min	Percentage of cells with *n* foci				
		1	**2**	**3**	**⩾4**	**Diffuse**
E. coli HMS174(DE3)						
pAC-GFP/LtrA	25/50/180	12.1	65.5	19.6	1.4	1.4
	25/250/180	7.2	67.1	23.9	1.2	0.6
	30/50/150	12.8	65.3	20.7	0.4	0.8
	30/250/120	9.1	69.0	21.0	0.9	0.0
	37/50/90	16.2	60.6	19.4	1.5	2.3
	37/250/60	7.6	69.1	20.0	1.8	1.5
plac-GFP/LtrA	37/0/120	3.2	56.3	31.5	4.0	5.0
	37/250/120	3.0	60.2	34.8	0.8	1.2
pACD2X-GFP/LtrA	37/50/60	17.9	45.3	26.4	9.4	1.0
	37/100/60	3.6	47.2	41.1	6.7	1.4
	37/250/60	1.3	29.6	52.7	16.4	0.0
	37/500/60	1.9	30.7	46.2	20.3	0.9
	37/1,000/60	2.4	33.3	45.5	17.2	1.6
pACD2X-LtrA/GFP	37/250/60	2.0	19.0	44.0	35.0	0.0
pACDF-GFP/LtrA	37/250/60	2.4	40.6	34.7	22.3	0.0
pAC-GFP/RT	37/250/60	7.4	53.5	28.0	9.4	1.6
pAC-GFP/En	37/250/60	7.6	60.1	29.6	2.7	0.0
pAC-GFP/LtrA(2-200)	37/250/60	3.0	44.0	39.7	12.3	1.0
pAC-GFP/LtrA(201-400)	37/250/60	0.4	21.1	21.1	55.4	2.0
pBAD24-$icsA_{507\text{-}620}$::gfp^{*}	37/250/60	62.4	31.8	5.8	0	0
pBAD24-$icsA_{507\text{-}620}$::gfp^{*} + pACD2X-GFP/LtrA	37/250/60	14.6[†]	7.1	1.1	0	70.5
E. coli AQ10033(DE3) (*oriC*$^{+}$)						
pACSD2-GFP/LtrA	37/250/60	6.1	61.4	26.5	3.6	2.4
E. coli AQ10060(DE3) (*oriC*$^{-}$)[‡]						
pACSD2-GFP/LtrA	37/250/60	6.0	40.8	25.3	27.9	0.0
L. lactis NZ9800						
pLE-GFP/LtrA-RIG	30/25[§]/180	35/5[¶]	55	5	0	0

Table 2.1: Localization of GFP fusion proteins.

Indicated are the numbers of foci of LtrA-GFP fusion proteins detected by fluorescence microscopy. In cells with one or two foci, the foci were always at the poles. In cells with three foci, two were at the poles and the third was elsewhere. In cells with four or more foci, two were at the poles and the remainder were elsewhere. At least 200 cells were counted in each experiment and about 70 to 80% of cells were fluorescent. The percentages shown in table are for fluorescent cells.

[*]Data are for the localization of $IcsA_{507-620}$/GFP.

[†]Another 6.7% showed one fluorescent focus outside the pole area.

[‡]Of the AQ10060(DE3) *oriC*$^-$ cells, 27.9% formed filaments, most of which had more than four fluorescent foci.

[§]Value indicates 25 ng/ml nisin.

[¶]Values indicate that 35% of cells showed one focus apposed or near the membrane at a putative pole in elongated or linked cells, and 5% showed one focus elsewhere.

Table 2.2: Polar localization of GFP/LtrA in mutants and knockouts

Strains	Genes	Percentage of cells with n foci (%)				
		1	**2**	**3+**	**Diffuse**	**No Fluor.**
WM1996	WT	9.7	56.5	17.1	2.6	14.1
WM2001	*mreB*	34.6	26.5	5.6	2.1	31.2
HMS174(DE3)	*galU*	17.8	39.7	15.0	2.0	25.5
HMS174(DE3)	*rfe*	17.2	37.8	14.8	2.4	27.8

Cells expressing GFP/LtrA under the T7 promoter were grown at 37°C, and induced with 500 μM of IPTG at 30°C overnight. At least 200 cells were counted in each strain for fluorescence distribution pattern.

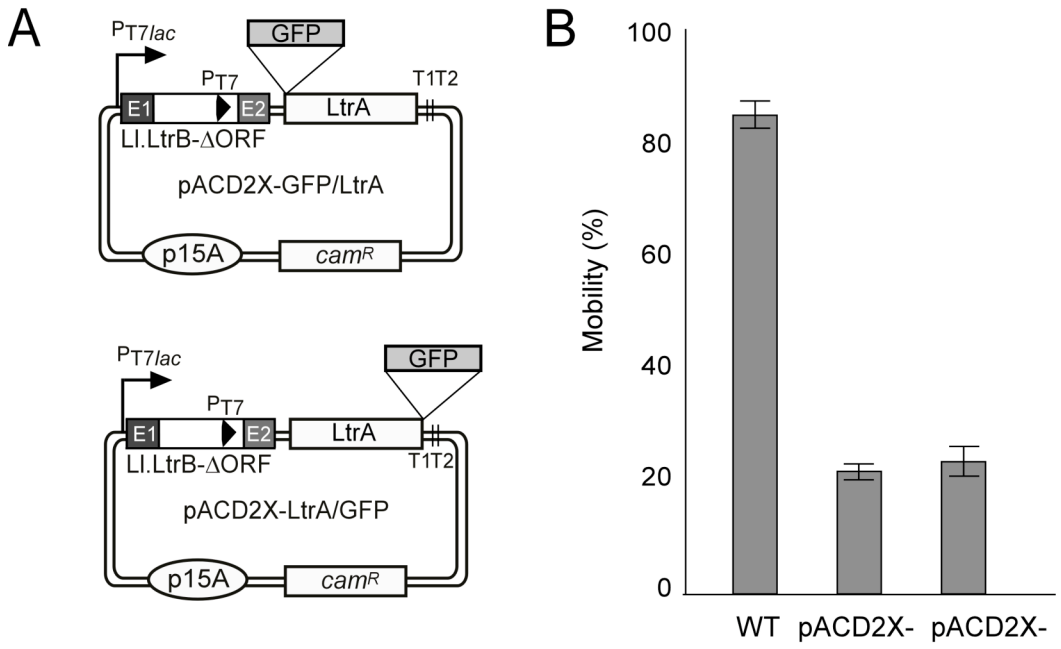

Figure 2.1: LtrA with GFP fusions are active in intron mobility.

(A) N-terminal (GFP/LtrA) or C-terminal (LtrA/GFP) fusion proteins were used in intron mobility assay (Figure 1.5 in Chapter 1). The Cam^R intron-donor plasmid pACD2X-GFP/LtrA or pACD2X-LtrA/GFP, which expresses intron and LtrA with GFP fusion protein under a phage T7 promoter, is cotransformed into *E. coli* HMS174(DE3) with the Amp^R recipient plasmid pBRR3-ltrB. The latter contains the Ll.LtrB intron target site upstream of a promoterless tet^R gene. Insertion of the intron carrying the T7 promoter into the target site activates the tet^R gene, and mobility frequencies are measured from the ratio of $(Tet^R+Amp^R)/Amp^R$ colonies. (B) Mobility frequency of wild-type Ll.LtrBΔORF intron with LtrA (WT), N- or C- terminal GFP fusions to LtrA in RNPs. Error bars show the standard deviations for three independent experiments.

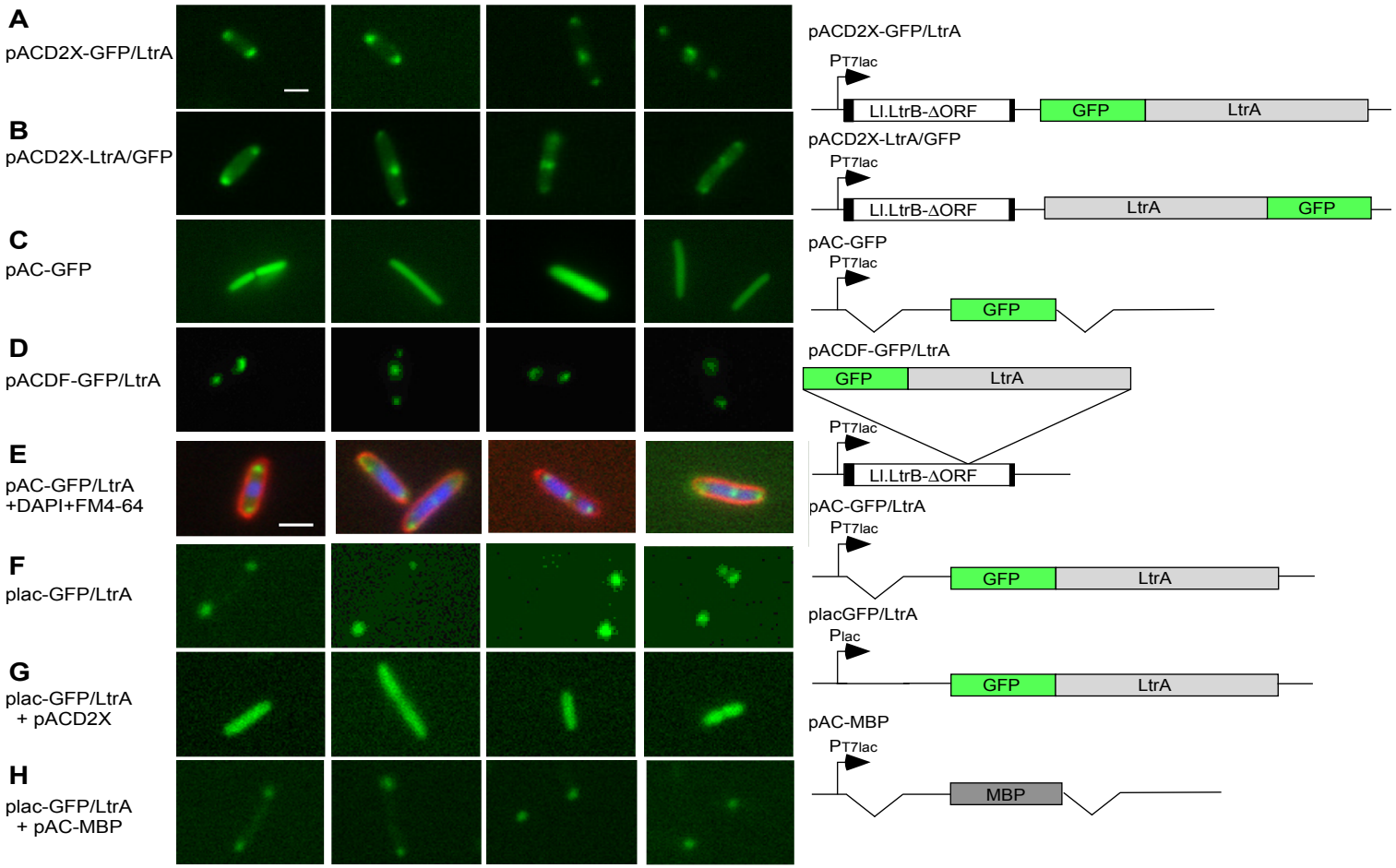

Figure 2.2: LtrA with GFP fusions are pole-localized in *E. coli*.

(A-H) Fluorescence microscopy. *E. coli* HMS174(DE3) containing the indicated plasmids was grown and induced with IPTG. In (E), DAPI and FM4-64 were added to stain DNA (blue) and cell membranes (red), respectively. Bar = 2 μm. Magnification x1.3 in panel E. Constructs are diagrammed to the right. The Ll.LtrB-ΔORF intron is indicated by an open rectangle with flanking exons shaded black. Deletions relative to the pACD2X-GFP/LtrA parent construct are indicated by breaks.

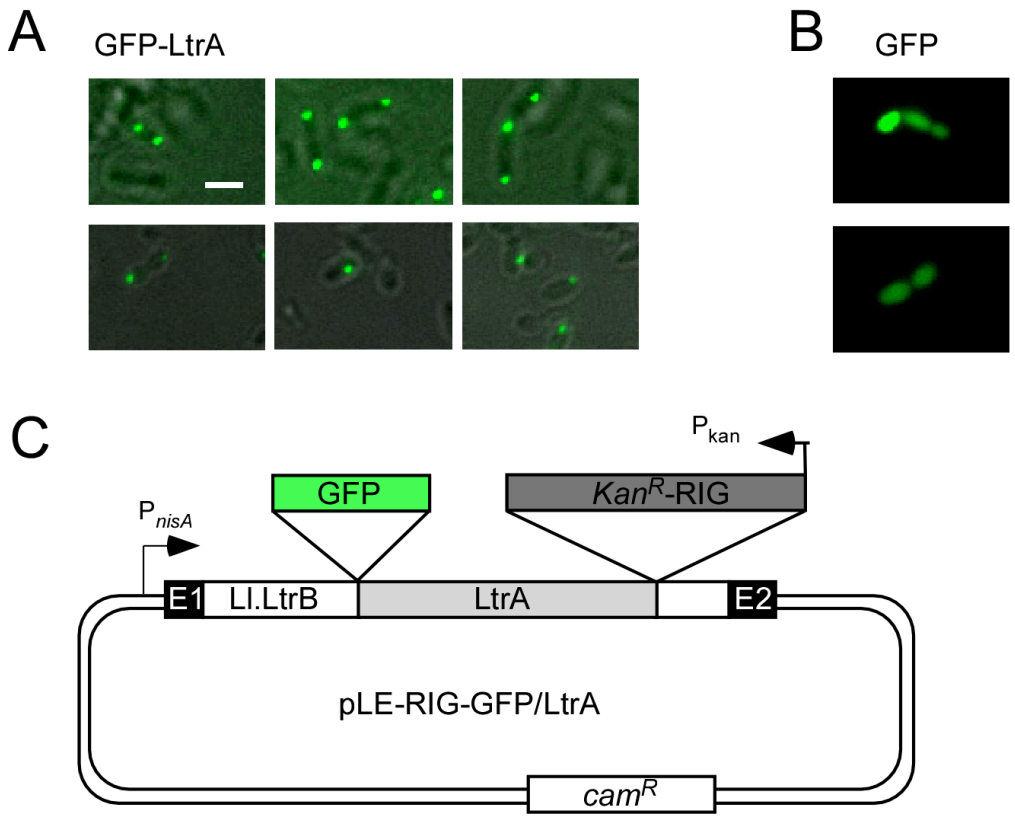

Figure 2.3: LtrA is pole-localized in *L. lactis*.

(A, B) Fluorescence microscopy. *L. lactis* NZ9800 strain containing pLE-RIG-GFP/LtrA to express N-terminal GFP/LtrA fusion (A) or pLE-RIG-GFP to express GFP alone (B), respectively, was grown and induced with nisin. In (A), fluorescence of GFP/LtrA fusion is superimposed over phase contrast images. Scale bar = 2 μm. (C) pLE-RIG-GFP/LtrA contains the Ll.LtrB intron and short flanking exons (E1, E2) cloned downstream of an inducible nisA promoter (P*nisA*). The ORF encoding the GFP/LtrA fusion is located in intron domain IV, just upstream of a Kan^R-RIG marker (Ichiyanagi *et al.*, 2002).

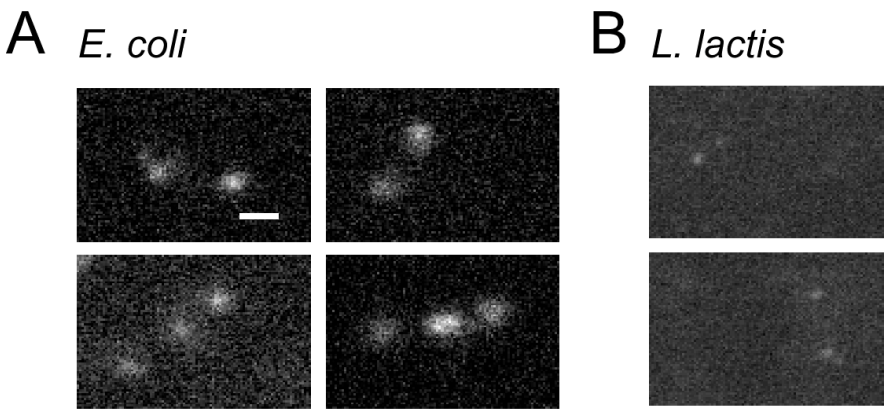

Figure 2.4: Immunofluorescence microscopy of LtrA.

(A) *E. coli* HMS174(DE3) containing pACD2X was induced with IPTG to express LtrA. (B) Endogenous LtrA was expressed from Ll.LtrB intron in *L. lactis* NZ9800. Both *E. coli* and *L. lactis* cells were fixed with a mixture of paraformaldehyde and glutaraldehyde. Immunofluorescence of LtrA was detected by anti-LtrA antibody and anti-rabbit IgG-FITC. Scale bar = 2 μm.

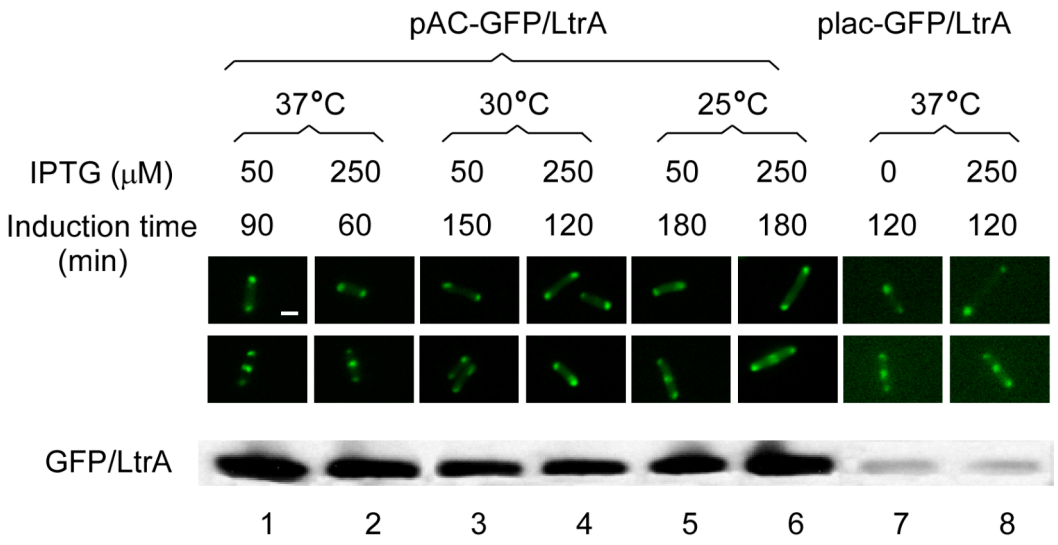

Figure 2.5: Localization of GFP/LtrA fusion protein at different expression levels.

Top, fluorescence microscopy. *E. coli* HMS174(DE3) containing pAC-GFP/LtrA or plac-GFP/LtrA was induced under the conditions indicated in the figure. Bottom, immunoblots. Total proteins isolated from the same culture used for fluorescence microscopy were probed with anti-LtrA antibody to detect the GFP/LtrA. Each lane was loaded with protein from the equal $O.D._{600}$ of cells, and equal loading was confirmed by Coomassie blue staining of a parallel gel (not shown). Scale bar = 2 μm.

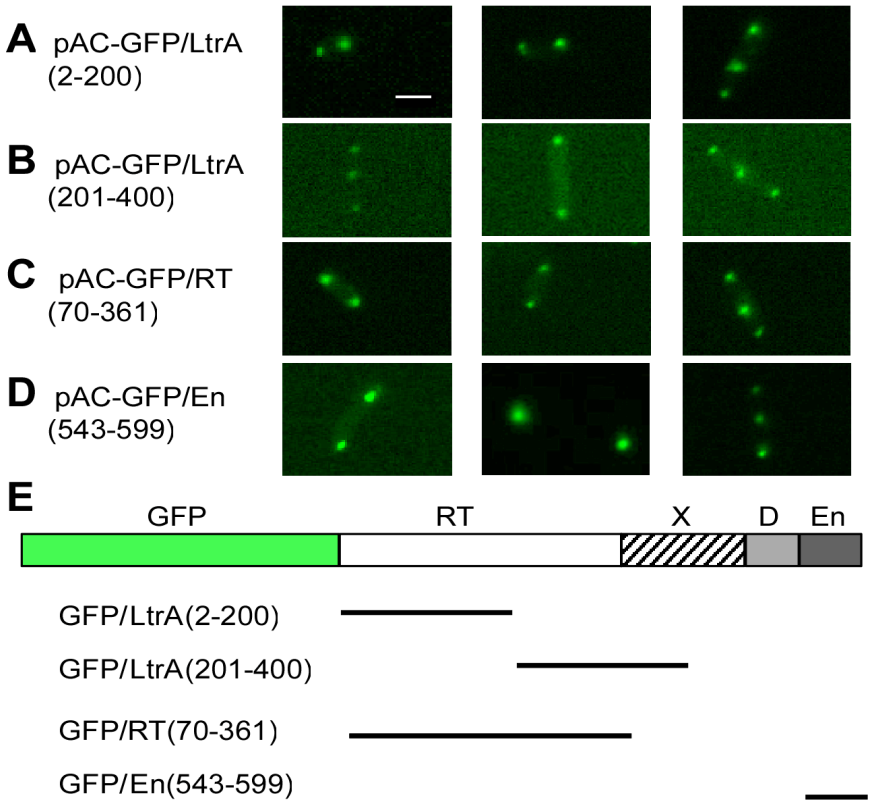

Figure 2.6: Localization of GFP fusions with different subsegments of LtrA.

(A-D) Fluorescence microscopy. *E. coli* HMS174(DE3) containing derivatives of pAC-GFP/LtrA with GFP fused to different subsegments of LtrA. Synthesis of the correct-sized protein was confirmed by SDS-PAGE and Coomassie blue staining (constructs A-C) or immunoblotting with anti-GFP antibody (construct D). Scale bar = 2 μm. (E) Schematic of the LtrA protein with segments used in the GFP fusions delineated below.

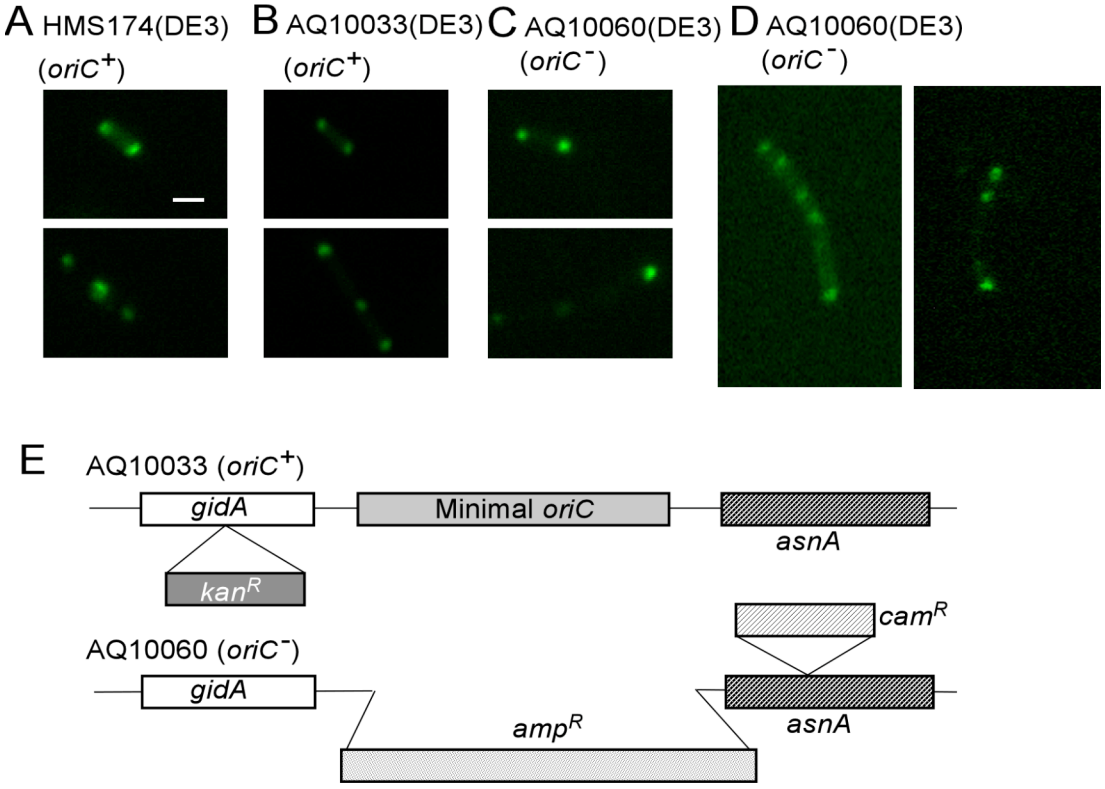

Figure 2.7: Polar localization of LtrA is not dependent upon *oriC* function.

(A-D) Fluorescence microscopy. *E. coli* HMS174(DE3) containing pACD2X-GFP/LtrA (*oriC*$^+$) (A), or AQ10033(DE3) *oriC*$^+$ (B), and AQ10060(DE3) *oriC*$^-$ (normal cells, C and filaments, D) containing pACSD-GFP/LtrA were induced with 250 μM IPTG for 1 h at 30°C. Scale bar = 2 μm. (E) Diagram of the *oriC* region in *E. coli* AQ10033 and AQ10060 in which the minimal *oriC* is replaced by *amp*^{R} gene.

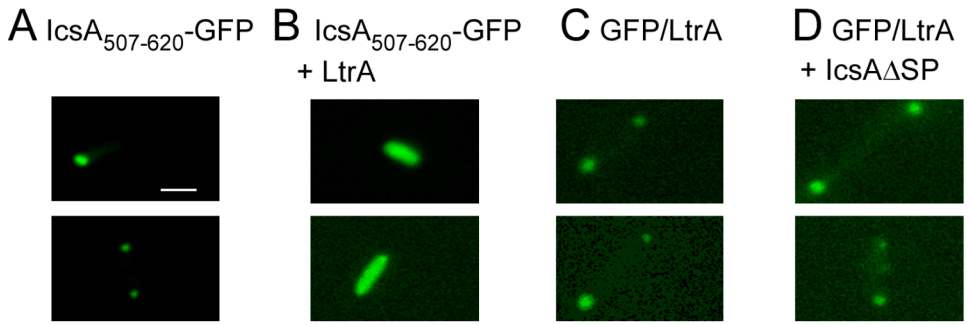

Figure 2.8: LtrA interferes polar localization of IcsA fragment.

(A-D) Competition experiments. *E. coli* HMS174(DE3) expressing (A) $IcsA_{507-620}$-GFP; (B) $IcsA_{507-620}$-GFP + LtrA; (C) GFP/LtrA; and (D) GFP/LtrA + IcsA with the signal peptide deleted. Cells were induced with 0.2% L-arabinose and/or 250 μM IPTG. Scale bar = 2 μm.

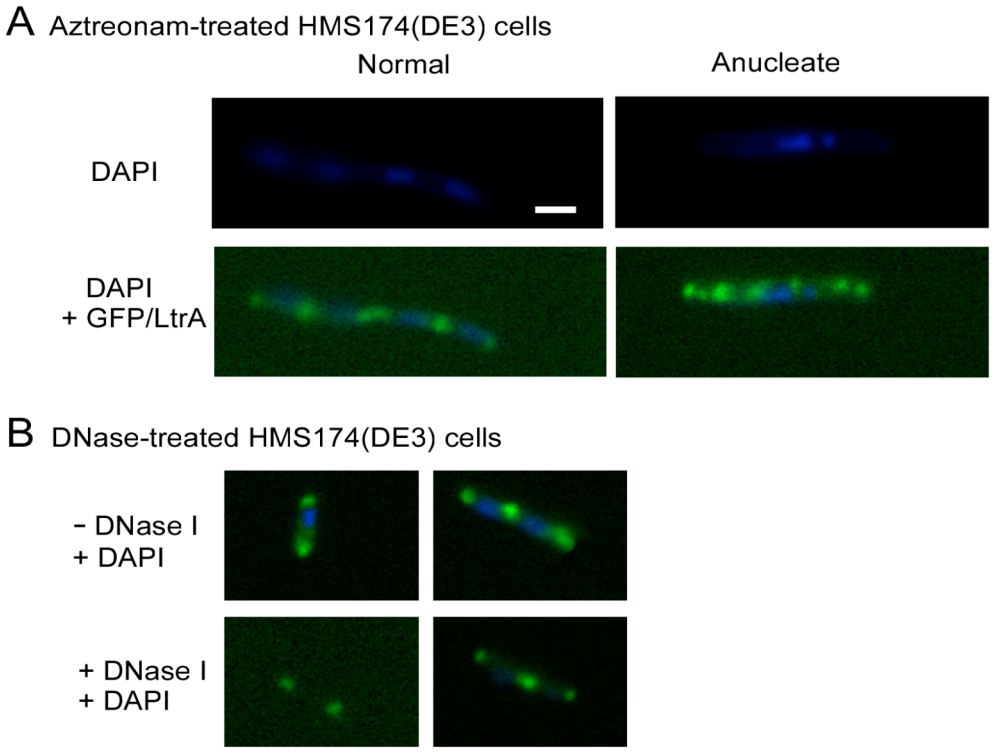

Figure 2.9: LtrA and IcsA may use related localization mechanisms.

(A) Effect of aztreonam. HMS174(DE3) expressing GFP/LtrA was induced with 250 μM IPTG in the presence of 1 μg/ml aztreonam. DAPI was added to stain DNA. (B) Effect of DNase I on GFP/LtrA localization. HMS174(DE3) expressing GFP/LtrA was induced and stained with DAPI. A portion of the cells was incubated with lysozyme (100 μg/ml, Sigma) and DNase I (100 units/ml, Invitrogen) for 1 h at room temperature, prior to fluorescence microscopy. Scale bar = 2 μm.

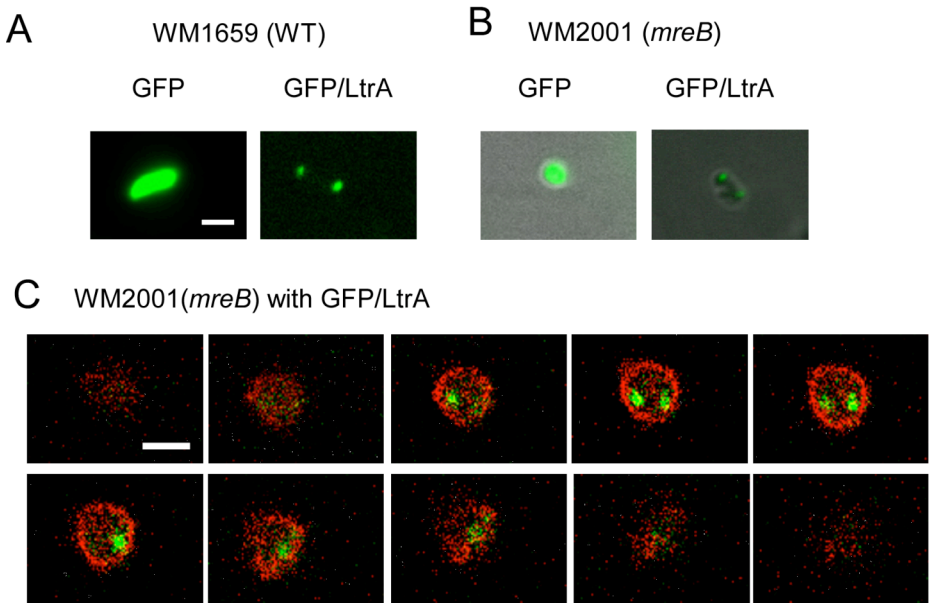

Figure 2.10: GFP/LtrA remains at poles in *mreB* mutant.

(A, B) Fluorescence microscopy. *E. coli mreB* mutant (WM2001) (B) and its parent strain WM1659 (A) express GFP/LtrA (pACD2X-GFP/LtrA) or GFP alone (pGFPuv). (C) Confocal fluorescence microscopy of GFP/LtrA (green) in *mreB* strain WM2001. The Cell membrane was stained with FM4-64 (red). Z series at 0.2 μm intervals. Scale bar = 2 μm.

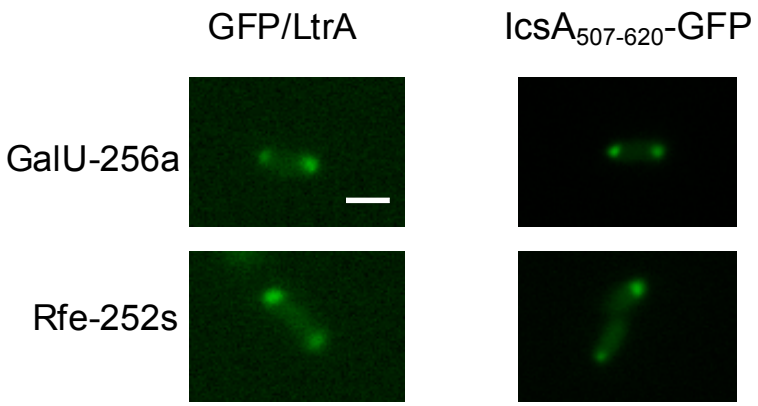

Figure 2.11: Polar localization of GFP/LtrA is not affected by mutations in LPS biosynthesis in *E. coli*.

E. coli genes *galU* and *rfe* were disrupted by Ll.LtrB-ΔORF targetrons (GalU-256a and Rfe-252s, respectively). Fluorescence of GFP/LtrA (left) or $IcsA_{507-620}$-GFP (right) expressed in these disruptants remained at cell poles. Scale bar = 2 μm.

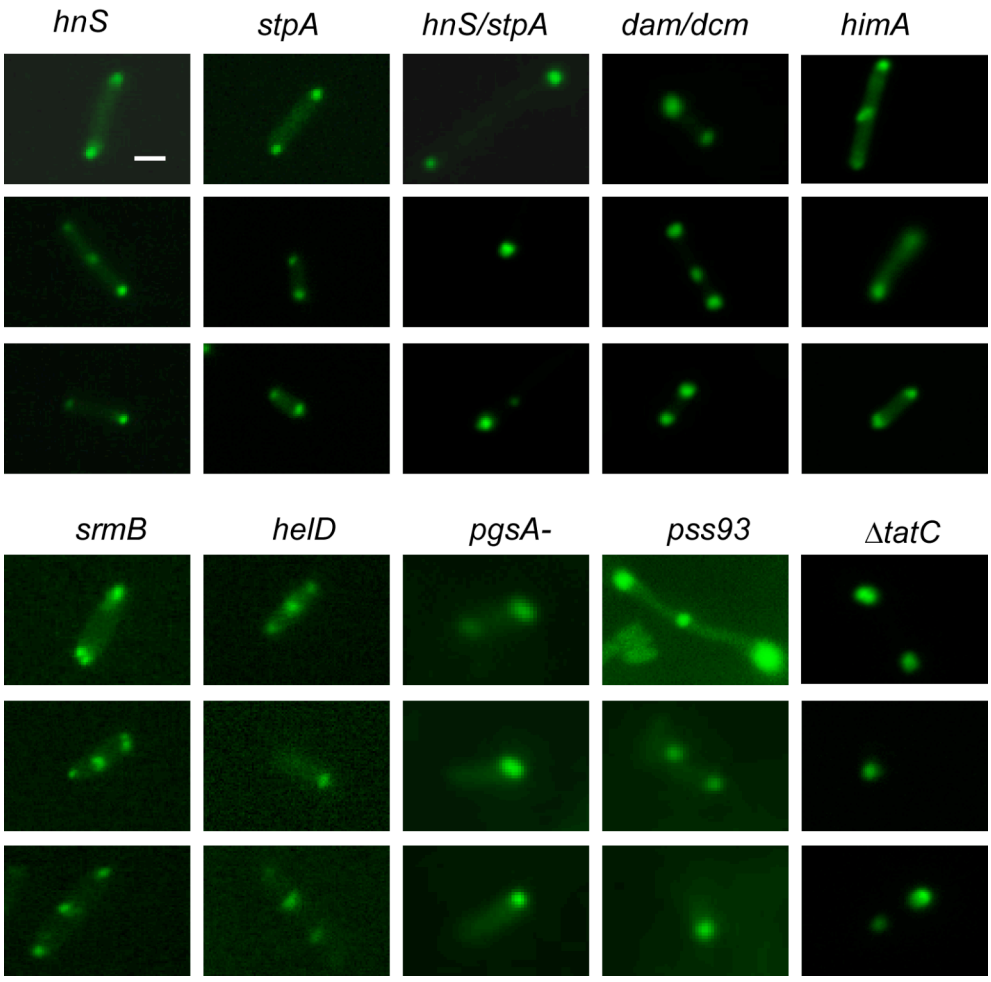

Figure 2.12: Polar localization of GFP/LtrA maintained in various mutant or knockout strains.

Fluorescence microscopy of GFP/LtrA fusion protein expressed under the T7 promoter from pACD2X-GFP/LtrA (*hns, stpA, hns/stpA, himA, srmB, helD*) or under the P*lac* promoter from plac-GFP/LtrA (*dam/dcm*, Δ*tatC*, *pgsA-*, *pss93*). Scale bar = 2 μm.

Chapter 3: Effect of host factors on LtrA localization and Ll.LtrB target site distribution

In the previous chapter, I showed by using LtrA/GFP fusions and immunofluorescence microscopy that the LtrA protein encoded by the Ll.LtrB group II intron is localized to the cellular poles in both *E. coli* and *L. lactis*. I showed further that, like *oriC*-linked sequences (Gordon *et al.,* 2001), GFP/LtrA remains pole localized in *oriC*$^-$ strains that use alternative replication origins. These findings suggest that LtrA is localized by interaction with cellular components that reside at the poles, and they are consistent with the possibility that the bipolar localization of LtrA contributes to the clustering of Ll.LtrB insertion sites in the *ori* and *ter* regions of the *E. coli* chromosome (Zhong *et al.,* 2003). To prove this connection, however, it is necessary to obtain mutations that alter LtrA's intracellular localization and show that they correspondingly change the distribution of Ll.LtrB insertion sites.

In this chapter, I carried out cell array screening using automated fluorescence microscopy to identify *E. coli* mutants with altered GFP/LtrA localization patterns. I isolated disruptants in five *E. coli* genes (*gppA*, *uhpT*, *wcaK*, *ynbC*, and *zntR*) that affect the bipolar localization of GFP/LtrA and lead to an increased proportion of cells with a more uniform cellular distribution of the protein. I then show that this altered GFP/LtrA localization pattern is correlated with a more uniform distribution of Ll.LtrB intron insertion sites in the *E. coli* genome. These findings support the hypothesis that the intracellular cellular localization of group II intron RNPs is a key determinant in the distribution of chromosomal insertions sites. The genes identified have seemingly disparate functions: *gppA* encodes guanosine pentaphosphatase; *uhpT* encodes a

membrane protein transporting hexose phosphate; *wcaK* encodes a predicted pyruvyl-transferase; *ynbC* encodes a protein of unknown function; and *zntR* encodes a transcriptional regulator that responds to elevated environmental metal concentration. However, a common feature is that in all of these disruptants as well as in a disruptant in the *ppx* gene which encodes exopolyphosphatase PPX, the altered GFP/LtrA localization pattern is correlated with a more dispersed distribution of intracellular polyphosphate. These findings suggest an unexpected connection between polyphosphate and protein localization in bacteria.

3.1 GENOME-WIDE SCREENING FOR HOST FACTORS AFFECTING LTRA LOCALIZATION

3.1.1 High throughput cell array screen for identifying *E. coli* mutants with altered LtrA localization patterns

To screen for host mutations affecting LtrA localization, I used the *mariner* transposon delivery vector pSC189 to generate an *E. coli* HMS174(DE3) library consisting of ~ 24,000 cells with *mariner* transposon insertions carrying a kan^R gene. The cell library was then transformed with the Cap^R intron-donor plasmid pACD2X-GFP/LtrA, which uses a T7*lac* promoter to express the Ll.LtrB-ΔORF intron with short flanking exons plus the GFP/LtrA fusion protein from a position just downstream of the 3' exon (Figure 2.1A in Chapter 2). The GFP/LtrA fusion protein splices the Ll.LtrB-ΔORF intron and remains associated with the excised intron lariat RNA in RNPs. I showed previously that RNPs containing the GFP/LtrA fusion are active in intron mobility and that the GFP/LtrA fusion is localized to the cellular poles in *E. coli* (Zhao & Lambowitz, 2005, Figure 2.2 in Chapter 2).

~9,600 Cap^R transformants were picked individually and stored in 96-well plates. These transformants were inoculated into new 96-well plates with fresh LB medium,

grown and induced with IPTG, as described in Materials and Methods, Chapter 4, then arrayed onto microscope slides with the help of Wei Niu (Marcotte lab, University of Texas at Austin), and screened by automated fluorescence microscopy (Figure 3.1). One batch of 4,900 cells was induced with 100 μM IPTG at 37°C, and a second batch of 4,700 cells was induced with 500 μM IPTG at 30°C. 76% of the transformants grew and showed detectable fluorescence in the arrays, with 0.08% of those induced at 37°C and 0.07% of those induced at 30°C showing reproducibly altered GFP/LtrA localization patterns (Figure 3.2).

Candidate strains with altered GFP/LtrA localization patterns in the primary screen were grown up from the original 96-well plate and examined by fluorescence microscopy to confirm the altered GFP/LtrA localization pattern. Of the initial 277 candidates, 36 strains showed similar altered fluorescence patterns in duplicate arrays (Table 3.1). Five out of these 36 candidates showed reproducibly altered GFP localization in fluorescence microscopy done manually with the original stock. The remainder did not show repeatable fluorescence distribution patterns and were not studied further.

3.1.2 Identification of *E. coli* genes affecting LtrA localization

The *mariner* transposon insertion sites in each of the five disruptant strains were identified by TAIL (thermal asymmetric interlaced) PCR (Liu & Whittier, 1995) and sequencing and found to be in the *wcaK, gppA, uhpT, ynbC,* and *zntR* genes. These insertion sites were confirmed by additional PCRs to amplify and sequence both the 5'- and 3'- transposon integration junctions (Figure 3.3) and by Southern hybridization of genomic DNA (Figure 3.4), which showed that each strain contains a single *mariner* transposon insertion at the expected site.

The *gppA* gene encodes guanosine pentaphosphatase (GPP), which is required for hydrolyzing pppGpp to the stringent response regulator ppGpp (Keasling *et al*., 1993); *uhpT* encodes a component of the hexose phosphate transport system (Hall & Maloney, 2001); *wcaK* is predicted to encode a pyruvyl-transferase in the colanic acid synthesis pathway (Stevenson *et al.,* 1996); *ynbC* encodes a 585-aa ORF of unknown function (Blattner *et al*., 1997); and *zntR* encodes the zinc-responsive transcriptional regulator, which controls the expression of the zinc export protein ZntA in response to high levels of environmental metals, including zinc, cadmium, and lead (Puskarova *et al.,* 2002, Newberry & Brennan, 2004). All of these are expressed as single genes, except for *uhpT*, which is the last gene in the four-gene *uhp* operon, and the *mariner* transposon is inserted at the 3' end of the *uhpT* gene. Thus, the phenotypes observed are likely due to disruptions of the gene in which the *mariner* transposon resides.

3.1.3 Intracellular localization of LtrA in disruptant strains

To further characterize the intracellular localization of GFP/LtrA in the disruptants, I examined 200 cells of each strain by fluorescence microscopy. In this analysis as well as that in the previous chapter, the wild-type strain expresses GFP/LtrA from the T7 promoter in plasmid pACD2X-GFP/LtrA. Immunoblots probed with anti-GFP antibody showed that the level of expression of the GFP/LtrA fusion protein in the disruptants was not elevated relative to that of the wild-type HMS174(DE3) (Figure 3.5), indicating that the diffuse fluorescence of GFP/LtrA is not due to over-expression of the protein. In the wild-type strain expressing GFP/LtrA under these conditions, about 20% of the cells showed no detectable fluorescence, and this was also true for each of the disruptants (Table 3.2). However, the proportion of cells showing no detectable

fluorescence was higher for some of the other strains and conditions described below, and for that reason, I have included this proportion in the Tables in this chapter.

At 30°C (Figure 3.2, Table 3.2A), each of the disruptants showed an increased proportion of cells with partially or completely diffuse GFP/LtrA fluorescence, (26.4-46.1% compared to 2.8% for wild type). The *gppA* and *wcaK* disruptants showed a high percentage of filamentous cells, which were excluded in the fluorescence pattern classification. The *gppA* disruptant showed the most pronounced phenotype with only 20.5% of the cells still showing polar localization, 39.5% showing partially or completely diffuse fluorescence (15.4 and 24.1%, respectively), and 16.9% filamentous, with irregular patches or dots of GFP fluorescence (Figure 3.2). The *wcaK* showed the least pronounced phenotype with 36.7% of the cells showing polar localization, and 26.4% of the cells showing partially or completely diffuse fluorescence. In the *gppA* and *zntR* disruptants, the predominant pattern was completely diffuse fluorescence (24.1 and 36.3%, respectively), whereas in the *uhpT* and *ynbC* disruptants, the predominant pattern was partially diffuse fluorescence (29.7 and 36.3%, respectively). The *wcaK* disruptant showed equal proportions of cells with partially and completely diffuse fluorescence (13.2%).

At 37°C (Table 3.2B), the disruptants showed similarly increased proportions of cells with partially or completely diffuse GFP/LtrA localization patterns (28.1-36.6%), although in most cases the proportion of cells showing the wild-type polar localization pattern appeared somewhat higher than at 30°C (37.1-46.7%). The *wcaK* disruptant now showed more cells with completely than partially diffuse fluorescence (19.8 and 11.0, respectively, and the *zntR* disruptant now showed more cells with partially diffuse than completely diffuse fluorescence (23.0 and 10.4%, respectively).

3.2 LtrA LOCALIZATION PATTERNS AFFECTED BY EXPRESSION LEVELS AND STRAIN SPECIFICITIES

3.2.1 LtrA localization patterns in *mariner* transposon disruptants with expression from the Pm promoter

Although immunoblots probed with an anti-GFP antibody showed that the level of expression of the GFP/LtrA fusion protein in the disruptants was not elevated relative to that of the wild-type HMS174(DE3) (see above Figure 3.5), it seemed desirable to additionally test whether the localization patterns in the disruptants might be affected by varying the expression level further using a different promoter. First, I tried to study the fluorescence distribution pattern of GFP/LtrA expressed under the *m*-toluic acid-inducible Pm promoter, which was found to be as effective as the T7 promoter for Ll.LtrB expression in intron-targeting experiment (Yao & Lambowitz, 2007). However, compared to the T7 promoter, fewer wild-type cells expressing GFP/LtrA under the Pm promoter showed detectable fluorescence (only 34% of HMS174(DE3) cells compared to 84% from the T7 promoter), and four of the five disruptants (*gppA*, *uhpT*, *ynbC*, and *zntR*) showed no detectable fluorescence above background. For the remaining disruptant, *wcaK*, only 21% of the cells showed detectable fluorescence of GFP/LtrA, which was predominantly polar localized (17%). Additionally, the GFP/LtrA fusion was detectable by immunoblotting only in the wild type and *wcaK* disruptant but not in the other disruptants (Figure 3.6). Because of the lower level of expression under the Pm promoter, it may be possible to detect GFP/LtrA only when it is clustered in discrete foci and not when it is partially or completely dispersed in the disruptants. In a control, I found that *m*-toluic acid, which is used to induce the Pm promoter, did not affect the fluorescence pattern of GFP/LtrA expressed under the T7 promoter (data not shown).

3.2.2 LtrA localization patterns in independently derived mutant strains

To further study the effect of mutations in the *gppA, uhpT, wcak, ynbC,* and *zntR* genes, I requested strains with complete deletions in these genes from the "Keio collection" maintained by Genobase (http://ecoli.naist.jp). The Keio strains (Baba *et al*., 2006) are an almost complete collection of *E. coli* deletions in which each gene is replaced with a kan^R gene in strain BW25113 ($lacI^q$ $rrnB_{T14}$ $\Delta lacZ_{WJ16}$ *hsdR514* $\Delta araBAD_{AH33}$ $\Delta rhaBAD_{LD78}$). Since the Keio collection does not include a *gppA* deletion, for this gene I requested a disruptant obtained with a Kan^R-miniTn10 transposon in strain KP7600 (F^- $lacI^q$ *lacZ*ΔM15 *galK2 galT22* Δλ *IN*(*rrnD-rrnE*)*1*) from Genobase. I confirmed that all of the requested strains have the expected deletion/insertion by PCR and sequencing the deletion junctions or integration junctions. To distinguish the mutant strains obtained from Genobase from those that I identified in the screen or constructed by targetron disruption, the Genobase strains are indicated by the gene name plus the letter (K) for Keio, *e.g., gppA*(K).

I studied GFP/LtrA localization in the Genobase strains by expressing GFP/LtrA under either the T7 promoter (T7 RNA polymerase gene introduced by phage λ DE3 lysogenization) or the Pm promoter from pBL1-based plasmids. With GFP/LtrA expressed under the T7 promoter, about 70-80% of the cells showed detectable fluorescence. However, only the *gppA*(K) strain showed an increased proportion of cells with partially or completely diffused fluorescence (21.9%), while the other four strains all displayed bipolar localization patterns (Figure 3.7, Table 3.3). Cell growth rates in these strains were similar to that of DH5α strain (doubling time = 50 min), except for a lower growth rate in the *gppA*(K) strain (doubling time = 80 min; not shown), possibly reflecting that the *gppA*(K) strain is derived form a different parental strain (KP7600) than the others (BW25113).

With GFP/LtrA expressed from the Pm promoter in pBL1, the percentages of fluorescent cells in the strains obtained from Genobase were comparable to those in wild-type DH5α strain (50-60% in knockouts and 65% in wild type, Table 3.3). However, all five of the disruptants obtained from Genobase showed predominantly polar localization of GFP/LtrA, and none showed a markedly increased proportion of cells with more diffused GFP/LtrA localization. As in HMS174(DE3), the expression level of GFP/LltrA under the Pm promoter was lower than that under the T7 promoter in the Keio strains (Figure 3.8). The different GFP/LtrA fluorescence distribution patterns in these strains might be due to the differences in strain background, protein expression levels, or both. Further, as noted previously, because of the lower expression level under the Pm promoter, GFP/LtrA fluorescence might not be detectable above background in cells in which the protein is delocalized.

3.3 Ll.LtrB intron mobility and chromosomal insertion site distribution in disruptants showing altered LtrA localization

3.3.1 Intron mobility in the disruptant strains

The results above show that the GFP/LtrA localization pattern is altered in five disruptants I identified by high-throughput cell array screening of the *E. coli* HMS174(DE3) disruption library, but that the protein localization pattern might be influenced by the GFP/LtrA expression level or strain background. Because the five strains I identified initially had reproducibly altered GFP/LtrA localization under the conditions of intron targeting experiment, they were still suitable for investigating how GFP/LtrA localization is correlated with the distribution of chromosomal Ll.LtrB insertion sites, which was a major objective of my research.

First, I determined mobility frequencies for the wild-type Ll.LtrB-ΔORF intron in each of the five *E. coli* HMS174(DE3)-derived disruptants. For these experiments, I used the standard *E. coli* plasmid assay in which an Ll.LtrB-ΔORF intron with a phage T7 promoter near its 3' end is expressed from a donor plasmid and integrates into a target site cloned upstream of a promoterless tet^R gene in a recipient plasmid, thereby activating the tet^R gene (Figure 1.5 in Chapter 1). The assays were carried out under two different induction conditions – 500 μM IPTG at 30°C for 3 h or 100 μM IPTG at 37°C for 1 h. Under both conditions, the *wcaK* disruptant showed substantially reduced intron mobility frequencies (26.9 and 18.1% of wild type, respectively), while the *gppA* strain showed substantially increased mobility frequencies (146.1 and 173.0% of wild type, respectively) (Figure 3.9, Table 3.4). The mobility frequencies in the remaining three disruptants (*uhpT*, *ynbC* and *zntR*) appear slightly elevated, but do not differ significantly from that of wild type (Figure 3.9, Table 3.4).

3.3.2 Chromosomal distribution of Ll.LtrB intron insertion sites in the disruptants

Next, I used the five disruptants to test whether the observed changes in LtrA localization pattern are correlated with an altered distribution of chromosomal Ll.LtrB-ΔORF intron-insertion sites. For these experiments, I used a pACD3-RAM library of Ll.LtrB-ΔORF introns with randomized target site recognition sequences (EBS1, EBS2, and δ) constructed by Zhong *et al.* (2003). The Tp^R-RAM marker carried in DIV of the intron consists of a small trimethoprim-resistance gene (tp^R) in the reverse orientation disrupted by the self-splicing *td* group I intron in the forward orientation. During retrotransposition of the Ll.LtrB intron via an RNA intermediate, the *td* intron is spliced generating an intact tp^R gene, which can then be selected after integration into a DNA target site (Figure 1.6 in Chapter 1).

To obtain chromosomal integrations, the library was transformed into wild-type HMS174(DE3) and each of the five disruptants. Cells were grown overnight, then inoculated 1:100 in LB medium, grown to $O.D._{600}$ = 0.3, and induced with IPTG. Cells in which the Ll.LtrB-ΔORF intron had integrated into chromosomal DNA were selected by trimethoprim resistance. In the wild-type strain, the frequency of Tp^R colonies was 1.1 x 10^{-7}. As expected from the relative mobility frequencies of the wild-type Ll.LtrB-ΔORF intron in these strains, the *gppA* disruptant had a higher frequency of Tp^R colonies (2.5 x 10^{-7}), the *wcaK* disruptant had a lower frequency (0.4 x 10^{-7}), and the remaining disruptants had about the same frequency (1 x 10^{-7}) as wild type. Thus the *gppA* and *wcaK* disruptants appear to have similar effects on mobility frequencies into plasmid or chromosomal target sites.

After selection, the Tp^R colonies were isolated and subjected to TAIL PCR to identify the intron-insertion sites. The insertion sites were amplified by using two nested specific primers complementary to the intron sequence together with a degenerate primer AD2 (Liu & Whittier, 1995), that anneals to chromosomal sequences upstream of insertion sites. The amplified bands from these samples were sequenced and the intron-insertion sites in the chromosome were identified. 50 or more target sites were identified for each strain, except in *zntR*, and the chromosomal distribution of these sites is plotted in Figure 3.10.

In agreement with the previous results (Zhong *et al.,* 2003), in the wild-type HMS174(DE3), 60% of the Ll.LtrB intron-insertion sites were located within 10% of the genome encompassing *oriC*. By combining these new chromosomal target site distribution data with those obtained previously (Zhong *et al.* 2003, Figure 1.6), I defined the preferentially targeted region around *oriC* in the wild-type strain to be between min 70 and 90 of the circular chromosome (21% of the genome, Figure 3.10). In the wild-type

strain, 79% of intron-insertion sites are located in this *ori* region encompassing *oriC* (Figure 3.10B) in this work and 75% in Zhong *et al.,* 2003 (Figure 3.10A).

By comparison to the wild-type strain, all the disruptants showed a more uniform distribution of Ll.LtrB insertions sites, with only 35-51% of the insertions in this *ori* region (Figure 3.10C-G, Table 3.4). Both *gppA* and *uhpT* strains showed only 46% (33/72, 30/65) of the insertion sites located in this region. Strain *wcaK* with the least diffuse GFP/LtrA fluorescence pattern (26.4%, Table 3.4) showed a higher proportion of insertions, 51% (51/100) in the *ori* region, but still lower than that in wild type. The percentage is even lower in the remaining two disruptants (*ynbC*, 38/100 =38%; and *zntR*, 9/26 = 35%). In each of two targeting experiments with the *zntR* disruptant, two insertion sites appeared repeatedly in the sequencing results (site 1325864, *btuR,* encoding cob(I) alamin adenolsyltransferase, and site 4199340, *zraP,* encoding Zn-binding periplasmic protein), likely reflecting that these insertions happened early in the induction process and were amplified by replication. This was observed only in the *zntR* disruptant, and the different repeated sites were counted once in the analysis. I was able to sequence only 26 different target sites in the *zntR* disruption.

3.4 EFFECT OF CHANGES IN POLYPHOSPHATE METABOLISM ON LTRA LOCALIZATION?

3.4.1 Does the *gppA* disruptant affect LtrA localization by decreasing ppGpp synthesis?

Both the *gppA* disruptant that I obtained by *mariner* transposon mutagenesis of HMS174(DE3) as well as the *gppA*(K) strain derived form strain KP7600 showed an increased proportion of cells with diffuse fluorescence of GFP/LtrA expressed under a T7 promoter. The enzyme guanosine pentaphosphatase (GPP), encoded by *gppA*, hydrolyzes pppGpp to ppGpp, known as the 'magic spot I' (Keasling *et al.,* 1993), and it also

processively hydrolyzes phosphoanhydride bonds of polyphosphate [poly(P)] chains to liberate orthophosphate (Keasling *et al.,* 1993).

In *E. coli*, the *relA* and *spoT* genes play a key role in the stringent response by determining the level of (p)ppGpp (Gentry & Cashel, 1996). RelA, the (p)ppGpp synthetase, synthesizes (p)ppGpp from ATP and GTP. SpoT is a bifunctional enzyme involved in ppGpp synthesis, like RelA, and also ppGpp hydrolysis, with its degradative activity outweighing its synthetic activity (Xiao *et al.,* 1991). *spoT* is an essential gene, whose knockout or deletion is lethal in a *relA*$^{+}$ background, but viable in *relA* knockout or mutant backgrounds. This double *relA*/*spoT* null strain CF1693 fails to accumulate ppGpp in stringent response (Xiao *et al.,* 1991).

Since GPP functions in hydrolysis of pppGpp to ppGpp, the *gppA* disruptant might contain a lower level of the 'magic spot I' ppGpp, or a higher level of pppGpp. To determine whether there is any relationship between LtrA localization and magic spot formation, I requested a *relA*/*spoT* mutant (AQ4319, *thr-1 leu-6 thi-1 lacY1 galK2 ara-14 xyl-5 mtl-1 proA2 his-4 argE3 rpsL-31 tsx-33 sup-37 thyA nfo-1::kan relA1 spoT thiA*) from Genobase, because the double disruption does not exist in the Keio collection. I observed polar localization of GFP/LtrA expressed under either the Pm or the T7 promoter (T7 RNA polymerase gene introduced by phage λ DE3 lysogenization) (Figure 3.12A). This result suggests that low ppGpp does not cause the altered GFP/LtrA localization pattern in the *gppA* disruptant. It remains possible, however, that elevated pppGpp levels in the *gppA* disruptant could affect GFP/LtrA localization. The high level of (p)ppGpp in cells also inhibits exopolyphosphatase – PPX activity and promotes poly(P) accumulation (Kuroda *et al.,* 1997).

3.4.2 Does the *gppA* disruptant affect LtrA localization by increasing poly(P) concentration?

In *E. coli*, poly(P) is produced from ATP by PPK (polyphosphate kinase, Akiyama *et al.,* 1992). The hydrolysis of poly(P) to orthophosphate *in vivo* is mainly accomplished by PPX, but with some contribution from GPP (Keasling *et al.,* 1993). The latter has extensive sequence identity (42.9%) to PPX (Kristensen *et al.,* 2004). *In vitro*, GPP has a K_m for binding its substrate polyphosphate much lower than that for pppGpp (Keasling *et al.,* 1993). Polyphosphate, the linear polymer chain of orthophosphate, exists in all organisms, but its physiological functions are largely unknown (Brown & Kornberg, 2004). Because of its high negative charge, *E. coli* poly(P) could potentially bind the positively charged LtrA protein (PI = 9.64) and affect its intracellular localization. Thus, it seemed possible that the *gppA* disruption affects LtrA localization by leading to accumulation of intracellular poly(P), which then binds and delocalizes LtrA.

To investigate whether changes in polyphosphate levels affect LtrA localization, I disrupted the *ppx* and *ppk* genes using Ll.LtrB targetrons (Ppk-1140a and Ppx-1051a, respectively). The disruptions were confirmed by PCR using primer pairs for the 5'- and 3'-junctions and sequencing of both the 5'- and 3'- junctions. Since the *ppx* gene follows *ppk* in the *ppk*/*ppx* operon (Akiyama *et al.,* 1993), there should be no activity of PPX in a *ppk* disruptant, while the *ppx* disruption should not affect PPK activity.

Next, I tested the localization of GFP/LtrA expressed under the T7 promoter in the disruptant strains (Figure 3.11A). In the *ppx* disruptant, which accumulates poly(P), I observed an increased proportion of cells with partial or completely diffuse fluorescence of GFP/LtrA (27.4%), although this proportion was not as high as in the *gppA* disruptant (39.5%). In the *ppk* disruptant, the GFP/LtrA remained at the poles as in wild type

(Figure 3.11A). Like most of the other disruptants derived from the wild-type HMS174(DE3), the level of GFP/LtrA expressed under the Pm promoter in targetron-induced *ppk* or *ppx* disruptants was too low to be detected by fluorescence microscopy (not shown).

Similar fluorescence distributions of GFP/LtrA expressed from a T7 promoter were observed with the *ppx* and *ppk* deletion stains obtained from Genobase – *i.e.*, an increased proportion of cells showing diffuse fluorescence of GFP/LtrA in Δ*ppx*(K)(DE3) (18.7%) but not in Δ*ppk*(K)(DE3) (T7 RNA polymerase gene introduced by phage λ DE3 lysogenization) (Figure 3.12A, Table 3.3). These results suggest that the elevated poly(P) concentration might contribute to diffuse localization of LtrA. However, the Δ*ppx*(K) strain expressing GFP/LtrA under the Pm promoter showed predominantly bipolar GFP/LtrA localization similar to wild-type strains (Figure 3.12A, Table 3.3). These findings suggest that both expression level and poly(P) may influence LtrA localization.

3.4.3 Ll.LtrB intron mobility frequency is unaffected in *ppk* and *ppx* disruptants

I also tested mobility of the wild-type Ll.LtrB-ΔORF intron in both sets of *ppk* and *ppx* disruptants (Figure 3.11B, 3.12B). Unlike the *gppA* disruption, which substantially increases the intron mobility frequency, the disruption or deletion of the *ppk* and *ppx* genes in either strain background had less of an effect. Mobility frequencies were equal to or slightly higher than wild type at either 30ºC (106% of the wild type in *ppk*, 93% in *ppx*; 95% in Δ*ppk*(K)(DE3), 109% in Δ*ppx*(K)(DE3)) or 37ºC (120% of the wild type in *ppk*, 103% in *ppx*; 98% in Δ*ppk*(K)(DE3), 113% in Δ*ppx*(K)(DE3)).

3.4.4 Fluorescence staining to visualize poly(P) granules in wild-type and disruptant strains

My findings for the *gppA* and *ppx* disruptants suggested that accumulation of poly(P) might be a factor that affects the intracellular localization of LtrA. When poly(P) binds DAPI, the emission wavelength of DAPI undergoes a red-shift with a maximum at 550 nm (yellow), compared to the 490 nm (blue) for DAPI bound to DNA (Huang *et al.,* 2005). Thus, poly(P) can be visualized by its yellow fluorescence in cells stained with DAPI. I used this method to visualize poly(P) in the wild-type and disruptant strains (Figure 3.13, Table 3.5). Because the filter in our fluorescence microscope allows any wavelength beyond 364 nm, poly(P) in the pictures appears yellow or orange. With 25 μg/ml of DAPI added during IPTG induction, about 25% of the cells showed poly(P) fluorescence in wild type and 48% in *gppA* disruptant cells (Table 3.5).

In the wild-type strain HMS174(DE3), poly(P) fluorescence was localized in discrete foci (mainly at poles 78.4%) in ~2/3 of the cells and diffuse in ~1/3 of the cells that showed detectable poly(P) fluorescence (Figure 3.13, Table 3.5). It was reported that poly(P) forms volutin granules normally localized at the poles in *Zoogloea ramigera* (Roinestad & Yall, 1970), so it could be contributing to the polar localization of LtrA and other proteins.

In contrast to the wild-type cells with polar localized poly(P), in the *gppA* disruptant cells showing poly(P) fluorescence, ~3/4 of the cells showed poly(P) dispersed over half or the entire cell (Figure 3.13), and only ~1/4 showing poly(P) in discrete foci (Table 3.5). As expected, the targetron-generated *ppx* disruptant, which was expected to lead to elevated poly(P), showed a higher proportion of the cells with detectable poly(P) fluorescence (59.8%), and an elevated proportion of cells showing diffused fluorescence (24.9%), although the latter was not as elevated in the *gppA* disruptant (37.4%, Table

3.5). In the *ppk* disruptant, which lacks polyphosphate kinase – PPK required for poly(P) synthesis – only a small proportion of the cells (6.9%) showed poly(P) fluorescence, also as expected (Figure 3.13, Table 3.5). The lower proportion of cells showing diffuse poly(P) fluorescence in the *ppx* disruptant than that in *gppA* disruptant may explain why the effect of the *ppx* disruption on GFP/LtrA localization is not as extreme as that of the *gppA* disruption. These findings suggest that although PPX may make a greater contribution to the hydrolysis of poly(P) than GPP (Keasling *et al.,* 1993), GPP may contribute more to its distribution in the cell.

Notably, the remaining four disruptant strains (*uhpT*, *wcaK*, *ynbC*, and *zntR*) from the cell array screen were similar to the *gppA* disruptant in showing a higher proportion of cells with detectable poly(P) fluorescence (47-55% compared to 25.4% for wild type) and a higher portion of cells showing diffused poly(P) fluorescence (from 15.8-23.6%, compared to 9% in wild type, Table 3.5). When stained with DAPI at the same concentration (25 μg/ml), a very low proportion of cells in strains requested from Genobase showed detectable poly(P) fluorescence (*gppA*(K), 3.4% is fluorescent, *ppx*(K), 7.0%), which may explain why these strain display mainly polar localization of GFP/LtrA. These results suggest that the poly(P) accumulation and distribution in cells may affect protein localization patterns, as is observed with GFP/LtrA.

Finally, in the *gppA* disruptant, 78.4% of the cells showing poly(P) fluorescence displayed colocalization of GFP/LtrA fluorescence with poly(P) (Figure 3.14), while in the wild-type HMS174(DE3) about half (57.6%) showing colocalized GFP/LtrA in cells with poly(P) fluorescence. In the *ppx* disruptant, which displays a high proportion of cells showing focused poly(P) (34.9%, Table 3.5), the proportion of cells with GFP/LtrA and poly(P) colocalization was 45.5%. This colocalization suggests that the negatively charged poly(P) might bind to LtrA protein and affect its intercellular distribution.

3.4.5 Effect of poly(P) distribution on positively charged protein CYT-18 in disruptant strains

To test the possibility that poly(P) affects the intracellular distribution of positively charged proteins due to its negative charge, I examined the localization of another positively charged protein – CYT-18 (PI = 9.29) in the same series of disruptants that affected LtrA localization. CYT-18, the *N. crassa* mitochondrial tyrosyl-tRNA synthetase, shows polar localization when expressed as a GFP fusion in wild-type HMS174(DE3) (Zhao & Lambowitz, 2005, Chapter 2). In five of the disruptant strains with more diffuse distribution of poly(P) (*gppA, uhpT, ynbC, zntR,* and *ppx*), the GFP/CYT-18 fusion protein shows more diffuse fluorescence similar to GFP/LtrA (Figure 3.15). In the *wcaK* disruptant strain, however, the distribution of GFP/CYT-18 fluorescence is less diffuse and similar to wild-type polar localization in the majority of cells (Figure 3.15). The *wcaK* disruptant strain also shows the most focused distribution of poly(P) (32.4% of the cell, Table 3.5), which is very close to that in the *ppk* disruptant strain (34.9%, Table 3.5), which shows the wild-type polar localization of GFP/CYT-18 (Figure 3.15). Thus, the more diffuse distribution of LtrA but not CYT-18 in the *wcaK* disruptant may reflect that LtrA is more sensitive than CYT-18 to changes in the intracellular distribution of poly(P). It is also possible that the *wcaK* disruptant has additional more direct effects on the localization of LtrA that differ from those on CYT-18. Together, these results suggest that the negatively charged poly(P) might bind to LtrA and other positively charged proteins and thus affect their intercellular distribution.

3.5 DISCUSSION

3.5.1 Host factors affecting LtrA localization

From cell array screening experiments, I identified five disruptants (*gppA, uhpT, wcaK, ynbC,* and *zntR*) in which a higher proportion of cells (30-40%) show more diffuse

fluorescence of GFP/LtrA compared to wild type (2.8%). *gppA* encodes GPP, which functions in cell stringent response by producing ppGpp, and also in hydrolysis of polyphosphate to orthophosphate. *uhpT* is involved in the hexose phosphate transport system. WcaK is predicted to be a pyruvyl-transferase in colanic acid synthesis. *ynbC* encodes a protein of unknown function. *zntR* encodes a transcriptional regulator responding to environmental metal concentration. Fluorescence microscopy of these disruptants revealed a common characteristic that these strains – the accumulation and more diffuse distribution of intracellular poly(P). Poly(P) is known to accumulate in response to a number of different cellular stresses (Rao *et al.,* 1998), and I suggest that different cellular stresses are a common feature that results in accumulation of poly(P) in the different disruptants. Together, these findings suggest that the negatively charged poly(P) may bind to the positively charged LtrA protein and affect its intracellular localization, with the more diffuse distribution of poly(P) in the disruptants delocalizing LtrA form the cellular poles.

3.5.2 LtrA localization contribution to intron integration

Consistent with the more diffuse fluorescence distribution of GFP/LtrA in these disruptants, chromosomal targeting experiments in these strains showed a less pronounced target site preference for integration of the of Ll.LtrB intron around *oriC*, compared to that in wild-type strain. These findings indicate that the intracellular localization of GFP/LtrA is correlated with intron integration site preference. Compared to the 79% of the target sites located in min 70-90 of the genome around the origin region in wild type, only 35-51% of the target sites are in this region in the five disruptants (Figure 3.10, Table 3.4). All five disruptants showed a more diffuse distribution of GFP/LtrA fluorescence in 30-40% of the cells. The fact that not all the cells are affected

might explain why I did not observe completely uniform distribution of intron integration sites. Excluding the *zntR* disruptant, for which I sequenced only 26 different target sites, the most uniform target site distribution was observed in the disruptant in the *ynbC* gene (38% in the *ori* region), which encodes a protein of unknown function.

In the five disruptants displaying diffuse GFP/LtrA fluorescence, if we combine the proportions showing partial and completely diffuse fluorescence, we notice that disruptant *wcaK*, containing the lowest percentage of cells showing diffuse fluorescence of GFP/LtrA (26.4%), also has the highest *ori* region preference of intron target site distribution in the genome (51%, Table 3.4). Similarly, the disruptant *ynbC*, which shows the highest proportion of diffuse fluorescence of GFP/LtrA (46.1%), displays the least clustering of intron target site distribution (38%, Table 3.4, Figure 3.10). These results further support the hypothesis that physical distribution of LtrA in cells affects the accessibility of RNPs to different chromosome regions and thus the target sites preference.

3.5.3 GFP/LtrA localization pattern is affected by protein expression level in deletion strains

The GFP/LtrA fusion protein expressed under the T7 promoter showed diffuse fluorescence in the five disruptant strains generated by insertion of *mariner* transposon in wild-type HMS174(DE3) cells. In another *gppA* disruptant with a miniTn10 insertion obtained from Genobase (*gppA*(K)), 21.9% of the cells showed a diffuse distribution of GFP/LtrA fluorescence when expressed under the T7 promoter. However, in the deletion strains (Δ*uhpT*(K), Δ*wcaK*(K), Δ*ynbC*(K), and Δ*zntR*(K)) corresponding to the other four disruptants identified by screening, GFP/LtrA showed the wild-type polar localization pattern. When expressed at lower level under Pm promoter, GFP/LtrA showed the typical polar localization distribution in all the strains from Genobase, including *gppA*(K).

Although the Pm promoter was shown to be as effective as the T7 promoter in intron targeting (Yang & Lambowitz, 2007), the expression level of GFP/LltrA under the Pm promoter was much lower than that under the T7 promoter in the disruptants in HMS174(DE3) background, with protein expression and fluorescence detectable only in the *wcaK* disruptant and wild-type HMS174(DE3). This inconsistency suggested that GFP/LtrA distribution could be affected by protein expression level and/or genetic background.

3.5.4 Poly(P) effect on LtrA localization

The linear polymer chain of orthophosphate exists in all organisms, but its physiological functions are largely unknown (Brown & Kornberg, 2004). It has been suggested to be an energy source and ATP substitute, chelator of metal ions, regulator for stress, survival, and development (reviewed in Kornberg, 1995). The existence of poly(P) in membrane calcium channels (Reusch *et al.,* 1997) suggests a contribution to the structure and function of membrane channels. Roinestad and Yall (1997) reported that poly(P) granules are usually localized at poles in *Z. ramigera*. In addition, studies showed that poly(P) may have a structural organization similar to nucleoid and that positively charged histone-like proteins may have a high affinity for poly(P) (Brown & Kornberg, 2004). The latter finding raises the possibility that poly(P) exposes condensed parts of the genome and results in a more uniform distribution of intron insertion sites. The negatively charged poly(P) may bind the positively charged LtrA and influence its location in *E. coli* cells.

Since the *E. coli* PPX and GPP proteins have extensive sequence identity and both proteins function in poly(P) degradation (Krisensen *et al.,* 2004), I tested GFP/LtrA localization in *ppx* and *ppk* disruptants, which are the major enzymes involved in poly(P)

metabolism. In the *ppx* disruptant, which accumulates poly(P) in the cell, fluorescence microscopy of GFP/LtrA showed an increased proportion of cells with diffuse fluorescence patterns, similar to that in the *gppA* disruptant. By contrast, in the *ppk* disruptant, in which poly(P) synthesis is inhibited, GFP/LtrA still remains at the poles. These findings suggest that high poly(P) concentration is responsible for the more diffuse distribution of LtrA. Interestingly, the remaining four disruptants identified in cell array screening (*uhpT*, *wcaK*, *ynbC*, and *zntR*) also showed an elevated and more diffuse distribution of poly(P) fluorescence. Thus, all the disruptants may influence the GFP/LtrA distribution via their common effect on of poly(P) accumulation and distribution.

The high proportion of cells showing colocalization of poly(P) and GFP/LtrA is consistent with the hypothesis that the elevated poly(P) might interact with the basic LtrA protein electrostatically and delocalize LtrA from poles. The diffuse localization of another positively charged protein CYT-18 in *gppA*, *uhpT*, *ynbC*, *zntR*, and *ppx* disruptant strains further supports the possibility that poly(P) can influence the intracellular localization of positively charged proteins in cells. It is also possible that the defects in the disruptants affect membrane composition, cause the loss of sequestration ability at the pole area, and lead to aberrant localization of both proteins and poly(P).

3.5.5 Cell array method for genome-wide screening

Ever since the microarray method was developed using spotted nylon membranes in the early 1990s (Stoughton, 2005), this technique has been adapted and improved for different uses in biological studies, such as DNA microarray, protein arrays, tissue arrays, and cell arrays (Stoughton, 2005; Wheeler *et al.,* 2005; Coppola & Geshwind, 2006). By directly printing prokaryotic or eukaryotic cells onto microscope slides, cell arrays

greatly accelerate high-throughput screening for functional genetic studies (Wheeler *et al.,* 2005).

In this dissertation research, the combination of cell arrays and automatic microscopy enabled the screening of a large number of *E. coli* knockouts for aberrant protein localization. The automatic printing and photography of *E. coli* culture avoids the massive labor of manual fluorescence microscopy examination. With the help of Wei Niu (Dr. Edward Marcotte lab, University of Texas at Austin), I was able to finish two rounds of cell array printing and microscopy of 100 96-well plates of *E. coli* disruptants in five months, and the time for printing and photography was only two weeks.

Among 96,000 colonies screened, I selected an initial 277 candidates that showed altered GFP/LtrA localization in at least one array. After validating the fluorescence distribution data by two sets of microscopy photos and examining the individual candidate by fluorescence microscopy from the original stocks, I identified five disruptants that showed reproducibly altered GFP localization in fluorescence microscopy. Further experiments showed that these five disruptants not only affected GFP/LtrA localization in cells, but also decreased the clustering of Ll.LtrB integration sites in the *oriC* region, resulting in a more uniform distribution of integration sites throughout the *E. coli* genome. My screening results demonstrate the feasibility of using cell arrays and automated fluorescence microscopy for analyzing protein localization in bacteria, and suggest that cell arrays could also be used to study cell morphology.

There are some limitations on this automatic method. First, it is difficult to keep cell number constant in each photo. For high throughput screening, cells were cultured in 96-well plates, which requires that all the disruptants be inoculated and grow at the same time. Consequently, slow-growing disruptants have lower cell densities than other strains. Only about 10^{-3} μl of each culture was spotted onto the microscope slide. The microscopy

photographs are also limited by the visible field of a 100x microscope lens in order to obtain sufficiently high resolution to see localization differences. Thus, the total cell number in each microscopy photo could be 20-40 in a well growing culture and several or none in a poorly growing one. Second, although the autofocusing was preset for the detection of blue fluorescence of DAPI staining, a small percentage of culture photos were still not correctly focused. In our screening, about 3% of the photos could not be analyzed because of a focusing problem. I solved this problem by taking more than one set of data from duplicate printed microscope slides, because the chance that both sets of photos of a given culture were out of focus is very slim. By comparing the two data sets, we were able to overcome the focusing problem and occasional imperfect printing of spots. Finally, I note that the colony-picking step and manual microscopy photo screening were the two most time-consuming steps in our screening process. Automated colony picking and computer analysis of fluorescence distribution would further improve the method.

3.6 SUMMARY

The polar localization of LtrA in *E. coli* was a possible explanation for the clustering of Ll.LtrB integration sites in the *oriC* region of the *E. coli* chromosome observed by us and others (Zhao & Lambowitz, 2005; Beauregard *et al.,* 2006). In the screening of an *E. coli* disruptant library generated by *mariner* transposon mutagenesis, I used cell arrays and automated fluorescence microscopy to identify five disruptants (*gppA, uhpT, wcaK, ynbC, and zntR*) in which a substantially increased proportion of cells showed partially or completely diffuse GFP/LtrA fluorescence. Further, each of the strains showed a more uniform distribution of Ll.LtrB chromosomal integration sites than did the wild-type *E. coli* strain, showing that the distribution of chromosomal insertion

sites is correlated with LtrA's intracellular localization. Although the distribution of LtrA is influenced both by protein expression level and strain background, the consistency of LtrA localization and intron target site distribution under the same culture conditions provides a possible explanation for the observed integration site preference. Finally, analysis of polyphosphate distribution in the five disruptant strains (*gppA, uhpT, wcaK, ynbC,* and *zntR*) as well as the *ppk* and *ppx* disruptants suggested that LtrA localization is influenced by the intracellular accumulation of polyphosphate, with a more diffuse distribution of high level polyphosphate in the cell causing a more diffuse distribution of LtrA.

Table 3.1: Thirty-six candidates from cell array screening showed altered GFP/LtrA localization pattern from microscopic analysis of two duplicated printed arrays.

Screening	Candidates
1st, 37°C	10C2, 14D1, 23B9, 23E8, 24D8, 25D12, 25G11, 26A1, 26E8, 29A12, 30G2, 32H4, 33G1, 34A2, 37E5, 40C1, 43A12, 43B5, 43F3, 45A12, 49A3
2nd, 30°C	2f8, 7a3, 9a1, 13b12, 15e9, 16f4, 17c7, 18c10, 18d2, 22c6, 24g11, 28c7, 31c1, 34h12, 42h8

21 candidates from the first batch of 51 96-well plates screening with induction at 37°C are listed according to the plate number and location in the plate. 15 candidates from the second batch of 49 96-well plates screening with induction at 30°C are listed with lower case column numbers to distinguish candidates from the first batch.

Table 3.2: GFP/LtrA localization in wild-type HMS174(DE3) strain and disruptants identified in cell array screening.

Strains	GFP/LtrA localization pattern (%)				
	Polar	**Partial diffuse**	**Complete diffuse**	**Filaments**	**No Fluor.**
A. 30°C					
HMS174(DE3)	81.2	2.5	0.3	0.0	16.0
gppA	20.5	15.4	24.1	16.9	23.1
uhpT	32.5	29.7	13.2	1.1	23.5
wcaK	36.7	13.2	13.2	18.6	18.3
ynbC	27.5	28.2	17.9	0.0	26.4
zntR	35.7	9.7	36.3	0.7	17.6
B. 37°C					
HMS174(DE3)	82.7	0.6	0.3	0.0	16.4
gppA	40.6	11.3	16.8	11.0	20.3
uhpT	45.3	21.6	9.1	0.5	23.5
wcaK	37.1	11.0	19.8	15.1	17.0
ynbC	38.6	27.0	9.6	0.0	24.8
zntR	46.7	23.0	10.4	1.9	18.0

Cells expressing GFP/LtrA under the T7 promoter from pACD2X-GFP/LtrA were grown at 37°C, and induced with 500 μM IPTG at 30°C (A) or 100 μM IPTG at 37°C (B) overnight. At least 200 cells were counted in each strain for fluorescence distribution pattern.

Table 3.3: GFP/LtrA localization in strains obtained from Genobase.

	Strains	GFP/LtrA localization pattern (%)			
		Polar	**Partial diffuse**	**Complete diffuse**	**No fluor.**
A	HMS174(DE3)	81.2	2.5	0.3	16.0
	gppA(K)(DE3)	58.0	8.1	13.8	20.1
	Δ*uhpT*(K)(DE3)	71.9	1.8	2.8	23.5
	Δ*wcaK*(K)(DE3)	73.3	2.5	1.4	22.8
	Δ*ynbC*(K)(DE3)	69.4	1.7	2.3	26.6
	Δ*zntR*(K)(DE3)	71.9	2.4	2.2	23.5
	Δ*ppk*(K)(DE3)	83.8	1.6	2.2	22.4
	Δ*ppx*(K)(DE3)	60.6	10.6	8.1	20.7
	relA/spoT(K)(DE3)	69.2	2.0	4.3	24.5
B	DH5α	63.0	1.2	1.3	34.5
	gppA(K)	48.9	1.1	4.3	45.7
	Δ*uhpT*(K)	54.2	1.4	1.2	43.2
	Δ*wcaK*(K)	56.1	1.0	2.1	40.8
	Δ*ynbC*(K)	50.8	1.1	1.1	47.0
	Δ*zntR*(K)	43.4	0.6	4.1	51.9
	Δ*ppk*(K)	51.9	1.6	2.2	44.3
	Δ*ppx*(K)	51.1	1.3	1.4	46.2
	relA/spoT(K)	53.6	1.5	3.5	41.4

Cells expressing GFP/LtrA under the T7 promoter from pACD2X-GFP/LtrA (A) or the Pm promoter from pBL1-GFP/LtrA (B) were induced with 500 μM IPTG or 2 μM of *m*-toluic acid at 30°C overnight. Letter (K) after gene name indicates strains obtained from Genobase. More than 200 cells of each strain were counted.

Table 3.4: Ll.LtrB intron mobility and chromosomal targeting site distribution in wild-type HMS174(DE3), and disruptants from cell array screening.

Strains	**Intron mobility (%)**		**% of target sites in *ori* region***	**GFP/LtrA partial + complete diffuse (%)†**
	30°C	**37°C**		
HMS174(DE3)	100	100	79	2.8
gppA	173.0	146.1	46	39.5
uhpT	85.1	106.0	46	42.9
wcaK	18.1	26.9	51	26.4
ynbC	112.8	104.0	38	46.1
zntR	121.9	117.0	35	46.0

*The *ori* region is defined as the region of the *E. coli* choromosome from 70 to 90 min (21% genome) around replication origin, based on two sets of chromosome integration data in wild-type HMS174(DE3) and 100 target sites were analyzed in each strain (except *gppA*, 72 events, *uhpT*, 65 events, and *zntR*, 26 events).

†Percentage of cells showing partially and completely diffuse fluorescence of GFP/LtrA, calculated from data at 30°C in Table 3.2.

Table 3.5: Poly(P) distribution in disruptant strains.

Strains	**Poly(P) pattern (%)**		
	Focused	**Diffuse**	**No fluor.**
HMS174(DE3) (WT)	16.5	8.9	74.6
gppA	10.6	37.4	52.0
uhpT	28.0	15.8	56.2
wcaK	32.4	23.6	45.0
ynbC	19.9	23.4	56.7
zntR	22.7	19.0	58.3
ppk	6.6	0.3	93.1
ppx	34.9	24.9	40.2
gppA(K)	1.8	1.6	96.6
ppx(K)	1.0	6.0	93.0

Disruptants were generated by *mariner* transposons (*gppA, uhpT, wcaK, ynbC,* and *zntR* strains) or designed targetrons (*ppk* and *ppx* strains) from wild-type (WT) HMS174(DE3). *gppA*(K) and *ppx*(K) were obtained from Genobase. Overnight culture was stained with DAPI, 25 μg/ml. At least 200 cells were counted for each strain.

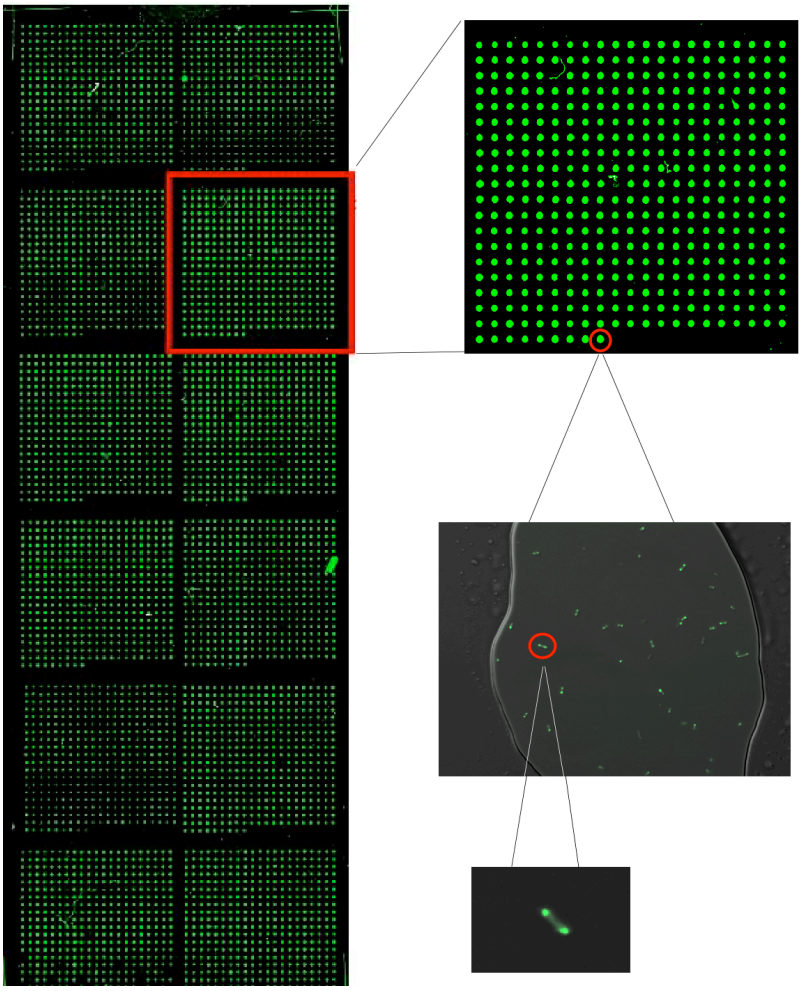

Figure 3.1: Cell arrays used to identify *E. coli* mutants affecting GFP/LtrA localization.

A library of *E. coli* HMS174(DE3) cells containing randomly inserted *mariner* transposons was transformed with pACD2X-GFP/LtrA to express a GFP/LtrA fusion protein. 9,600 Kan^R and Cam^R colonies were picked from LB plates into one hundred 96-well plates. In one experiment, cells in fifty one 96-well plates were induced with 100 μM IPTG at 37°C, and in another experiment, cells in the other forty nine 96-well plates were induced with 500 μM IPTG at 30°C. In both experiments, the ~5,000 colonies along with a wild-type control were arrayed on microscope slides and screened by automated fluorescence microscopy system to identify those having altered LtrA localization patterns. Fluorescence microscopy of a colony showing the wild-type pattern of bipolar localization of GFP/LtrA is shown as an example.

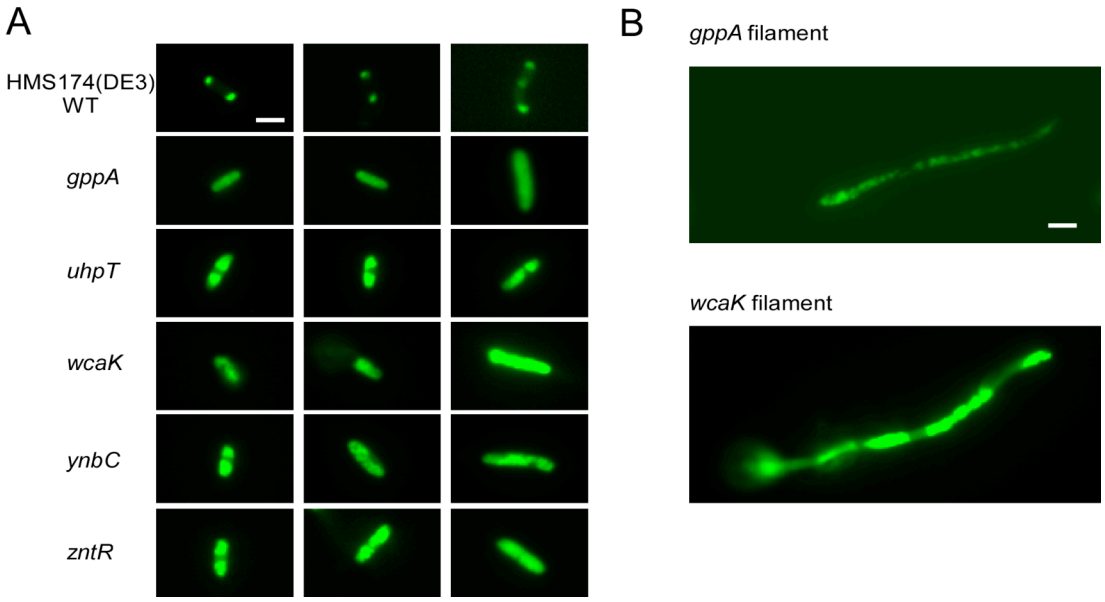

Figure 3.2: Fluorescence microscopy showing GFP/LtrA localization in wild-type HMS174(DE3) and disruptant strains.

Cells containing pACD2X-GFP/LtrA were induced with 500 μM IPTG at 30°C overnight. (A) shows wild-type HMS174(DE3) (WT) cells with a polar localization pattern and disruptant cells with completely or partially diffuse GFP/LtrA localization patterns. The localization patterns shown represent the most frequent pattern seen in each strain (Table 3.2). (B) GFP/LtrA localization in filamentous *gppA* and *wcaK* disruptants. Scale bar = 2 μm.

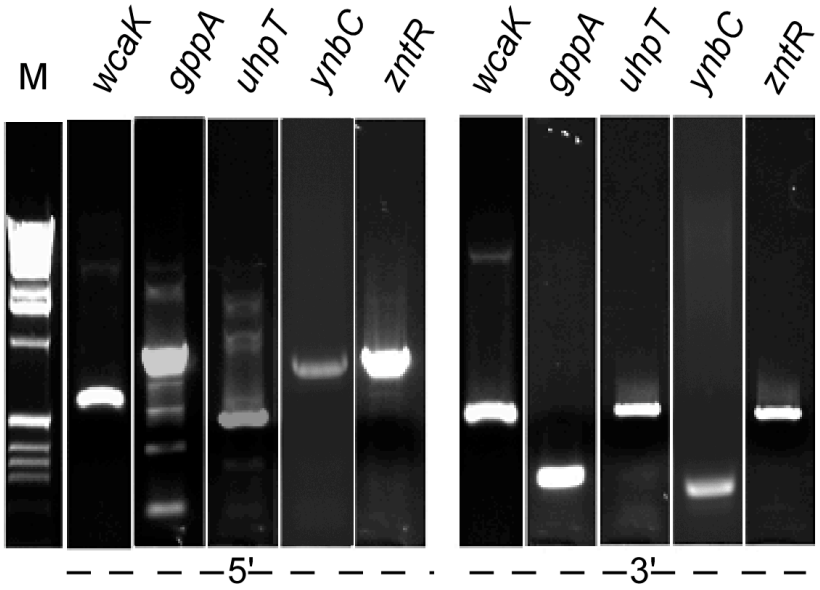

Figure 3.3: PCR amplification of *mariner* transposon insertion junctions in disruptant strains identified by cell array screening.

Genomic DNA from each disruptant was used as template to amplify the 5'- (left) and 3'- (right) junctions of the transposon insertion site, with one primer annealing to the predicted genomic DNA sequence and the other primer annealing near the 5' or 3' end of the transposon. The insertion junctions were confirmed by sequencing the PCR products (not shown). M, 1-kb DNA ladder (Invitrogen).

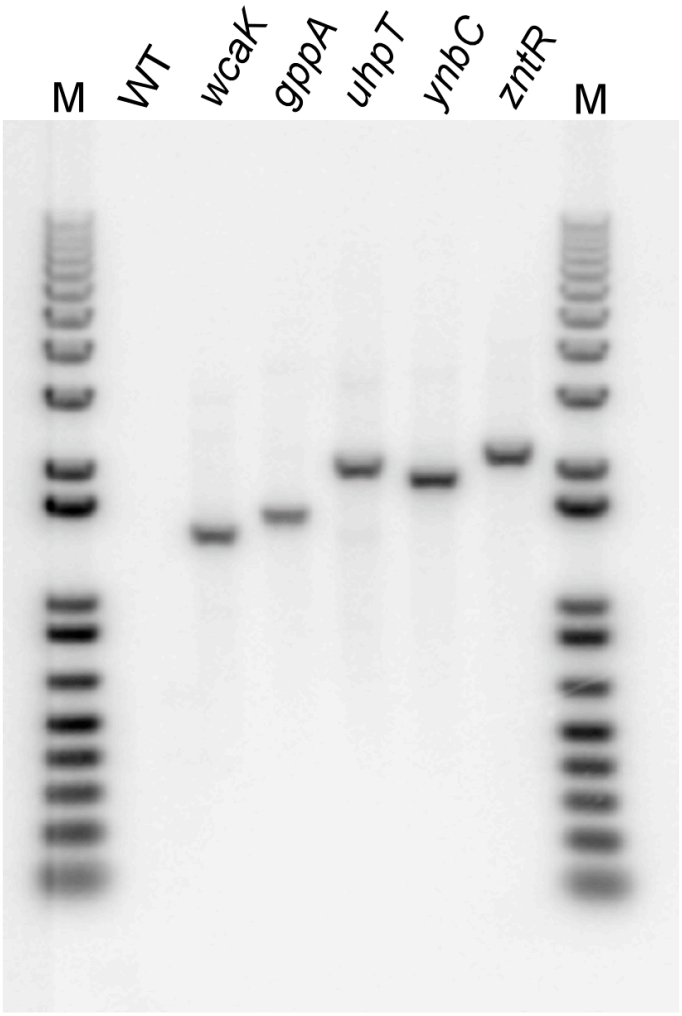

Figure 3.4: Southern hybridization of *mariner* transposons in disruptant strains.

Genomic DNA in the indicated disruptant strains with altered GFP/LtrA localization and wild-type HMS173(DE3) (WT) was used for Southern hybridization to verify single insertion of *mariner* transposon in the disruptants. In each sample, genomic DNA was digested by *Xcm*I, *Xma*I, and *Sac*II (New England Biolabs). The blot was probed with a ^{32}P-labeled PCR product corresponding to nts 1385-1868 of the *mariner* transposon. M, 1-kb plus DNA ladder (Invitrogen).

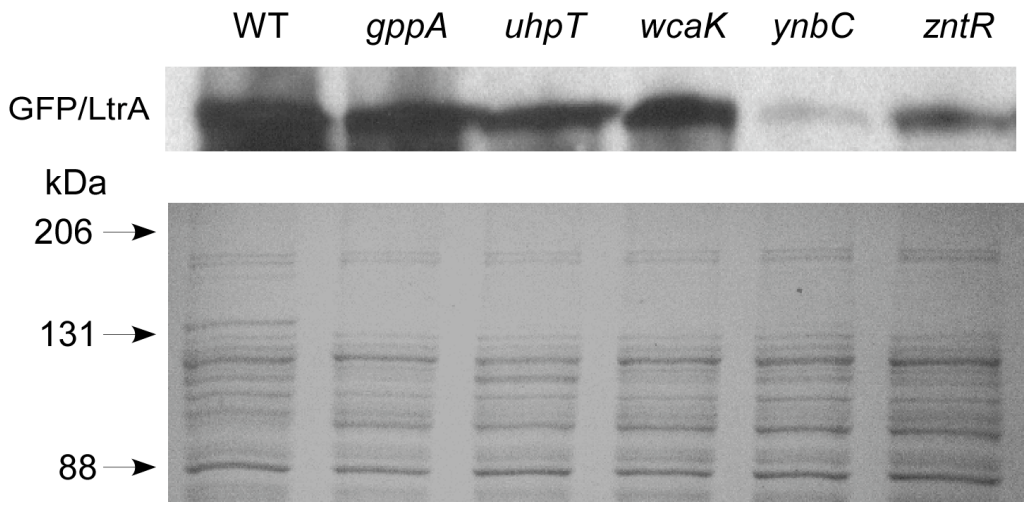

Figure 3.5: Immunoblotting of GFP/LtrA expressed under the T7 promoter in wild-type HMS174(DE3) (WT) and the indicated disruptant strains.

Top, immunoblot of GFP/LtrA fusion protein probed with anti-GFP antibody (JL-8, BD Biosciences). The immunoblot is over-exposed to show the light LtrA band in the *ynbC* disruptant. Bottom, parallel gel stained with Coomassie blue. Loading was normalized by $O.D._{600}$ of the induced cultures and Coomassie staining of a parallel gel. Arrows to the left of the gel indicate positions of size markers (Kaleidoscope Prestained Standard, BioRAD).

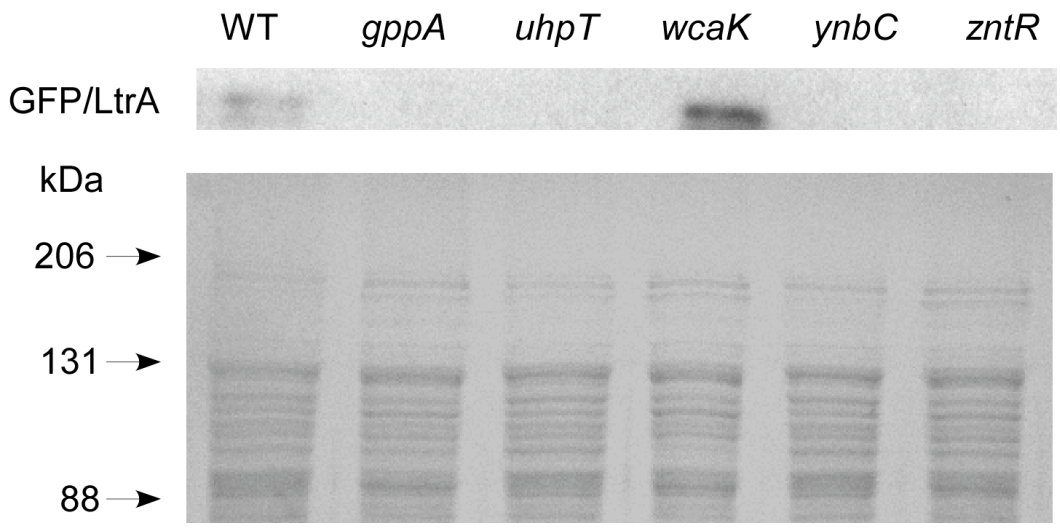

Figure 3.6: Immunoblotting of GFP/LtrA expressed under the Pm promoter in wild-type HMS174(DE3) (WT) and the indicated disruptant strains.

Top, immunoblot of GFP/LtrA fusion protein probed with anti-GFP antibody (JL-8, BD Biosciences). Bottom, parallel gel stained with Coomassie blue. Loading was normalized by $O.D._{600}$ of induced culture and Coomassie staining of a parallel gel. Arrows to the left of the gel indicate positions of size markers (Kaleidoscope Prestained Standard, BioRAD).

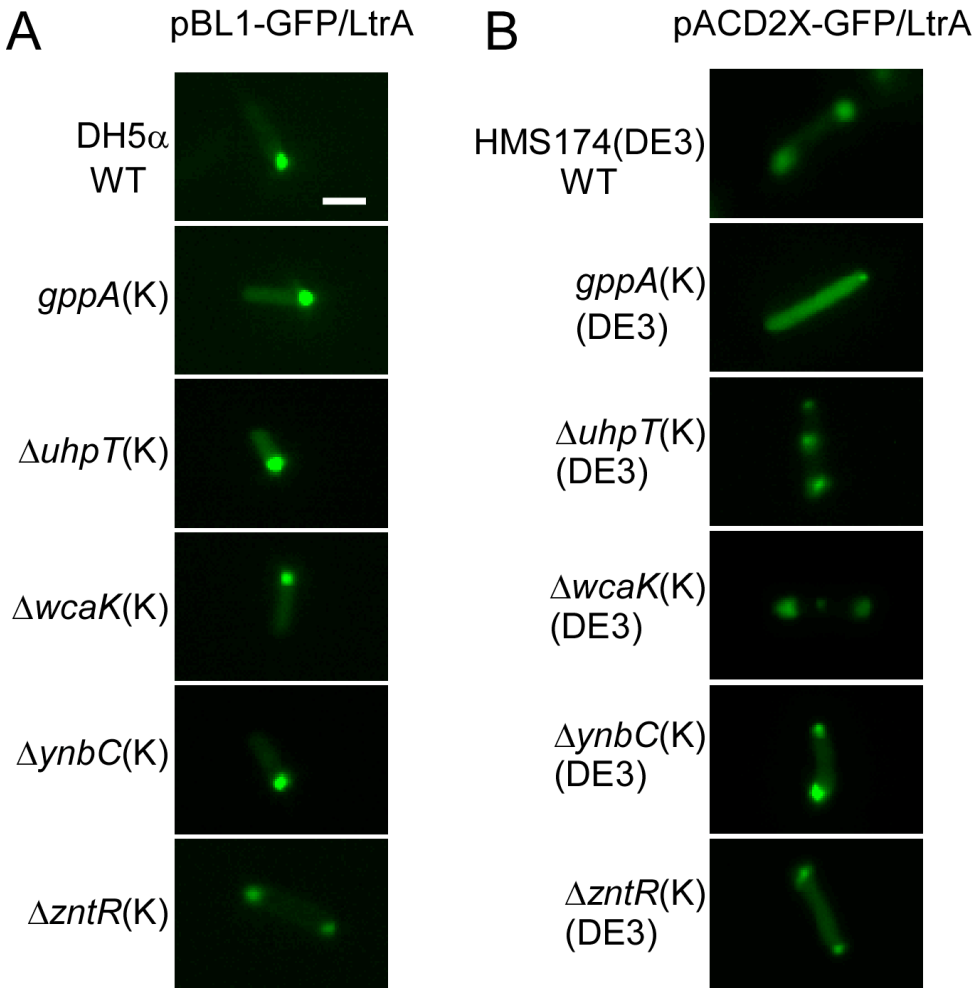

Figure 3.7: Fluorescence microscopy of GFP/LtrA in strains obtained from Genobase.

GFP/LtrA was expressed in wild-type *E. coli* DH5α, HMS174(DE3), or disruptant *gppA*(K) and Keio strains (*uhpT*(K), *wcaK*(K), *ynbC*(K), and *zntR*(K)) from Genobase (ecoli.naist.jp). The protein was expressed under the Pm promoter from pBL1-GFP/LtrA (A) or the T7 promoter from pACD2X-GFP/LtrA (B), in indicated strains with a T7 RNA polymerase gene introduced by λ(DE3) lysogenization. Letter (K) is used to indicate strains obtained from Genobase. The localization patterns represent the most frequent seen in each strain. Scale bar = 2 μm.

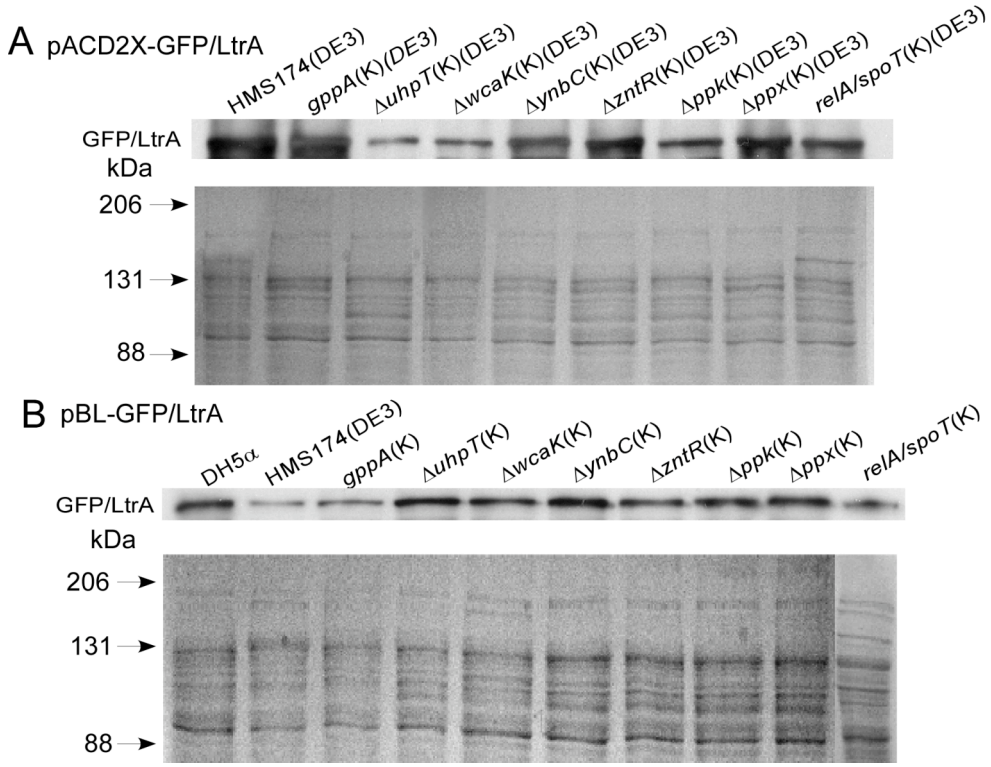

Figure 3.8: Immunoblotting of GFP/LtrA fusion protein in wild-type and knockout, mutant, or disruptant strains obtained from Genobase.

(A) GFP/LtrA was expressed under the T7 promoter from pACD2X-GFP/LtrA in wild-type HMS174(DE3), or strains from Genobase (disruptant *gppA*(K), or Keio collection Δ*uhpT*(K), Δ*wcaK*(K), Δ*ynbC*(K), Δ*zntR*(K), and mutant *relA*/*spoT*(K)).The strains were modified to carry a T7 RNA polymerase gene via λ(DE3) lysogenization. (B) GFP/LtrA was expressed under the Pm promoter from pBL1-GFP/LtrA in wild-type HMS174(DE3), DH5α, or indicated strains from Genobase. In (A, B), top, immunoblot of GFP/LtrA with anti-GFP antibody (JL-8, BD Biosciences); bottom, parallel gel stained with Coomassie blue. Loading was normalized by $O.D._{600}$ of induced culture and Coomassie staining of a parallel gel. Arrows to the left of the gel indicate positions of size markers (Kaleidoscope Prestained Standard, BioRAD).

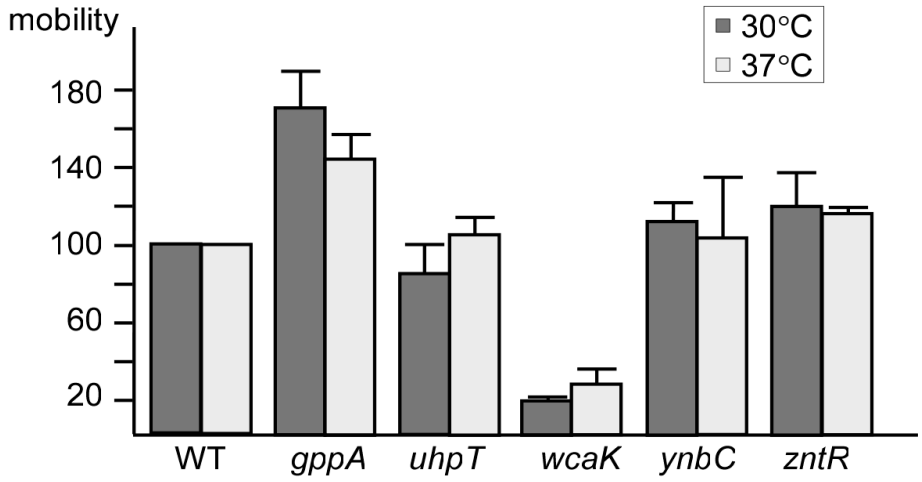

Figure 3.9: Intron mobility assays in wild-type HMS174(DE3) and the indicated disruptant strains.

E. coli plasmid assay for Ll.LtrB intron mobility was described in Chapter 1 (Figure 1.5). Mobility frequencies in wild-type HMS174(DE3) (WT) and disruptants *gppA*, *uhpT*, *wcaK*, *ynbC*, and *zntR*, were studied with the wild-type Ll.LtrB-ΔORF intron induced with 500 μM IPTG at 30°C or 100 μM IPTG at 37°C. Intron mobility frequencies were calculated as the ratio of Amp^R + Tet^R/Amp^R colonies, and were normalized to that of the wild-type strain assayed in parallel. (The mobility frequency in the wild-type strain was 70-80% in these experiments). Error bars indicate the standard deviations for three independent assays.

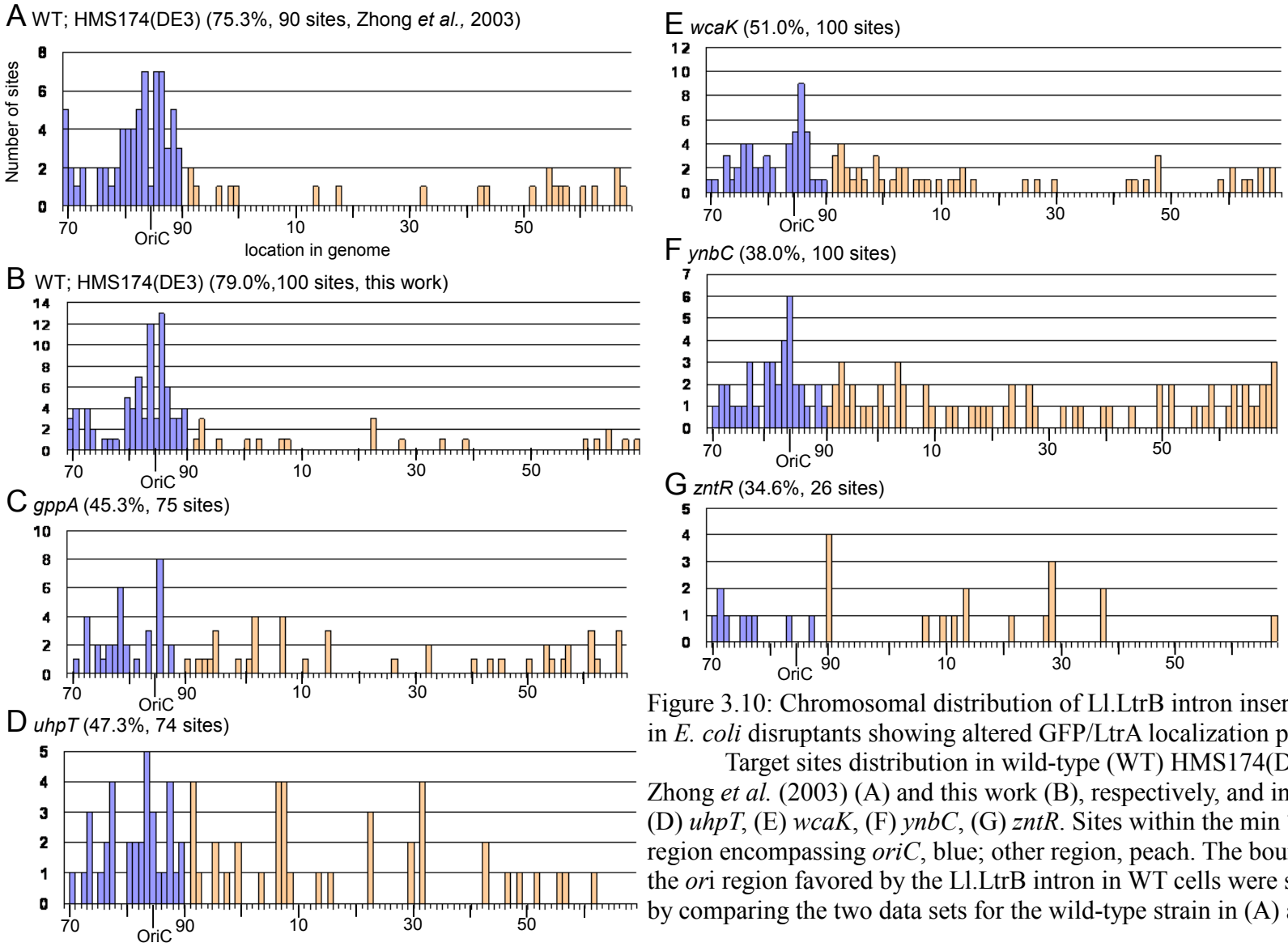

Figure 3.10: Chromosomal distribution of Ll.LtrB intron insertion sites in *E. coli* disruptants showing altered GFP/LtrA localization patterns.

Target sites distribution in wild-type (WT) HMS174(DE3) from Zhong *et al.* (2003) (A) and this work (B), respectively, and in (C) *gppA*, (D) *uhpT*, (E) *wcaK*, (F) *ynbC*, (G) *zntR*. Sites within the min 70-90 region encompassing *oriC*, blue; other region, peach. The boundaries of the *ori* region favored by the Ll.LtrB intron in WT cells were selected by comparing the two data sets for the wild-type strain in (A) and (B).

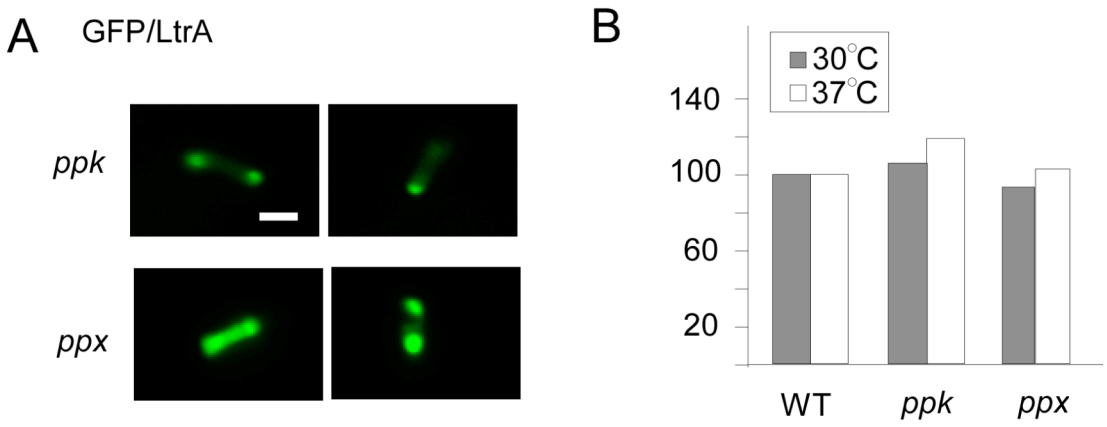

Figure 3.11: GFP/LtrA localization and intron mobility in *ppk* and *ppx* disruptants generated by using targetrons in wild-type HMS174(DE3) (WT).

(A) Fluorescence microscopy of GFP/LtrA expressed under the T7 promoter from pACD2X-GFP/LtrA in derivatives of HMS174(DE3) having disruptions in *ppk* or *ppx* gene. The disruptions were generated by site-specific insertion of targetrons (Ppk-1140a and Ppx-1051a). Scale bar = 2 μm. (B) Mobility of the wild-type Ll.LtrB-ΔORF intron in *ppk* and *ppx* disruptant strains at 30°C (gray bars) or 37°C (white bars).

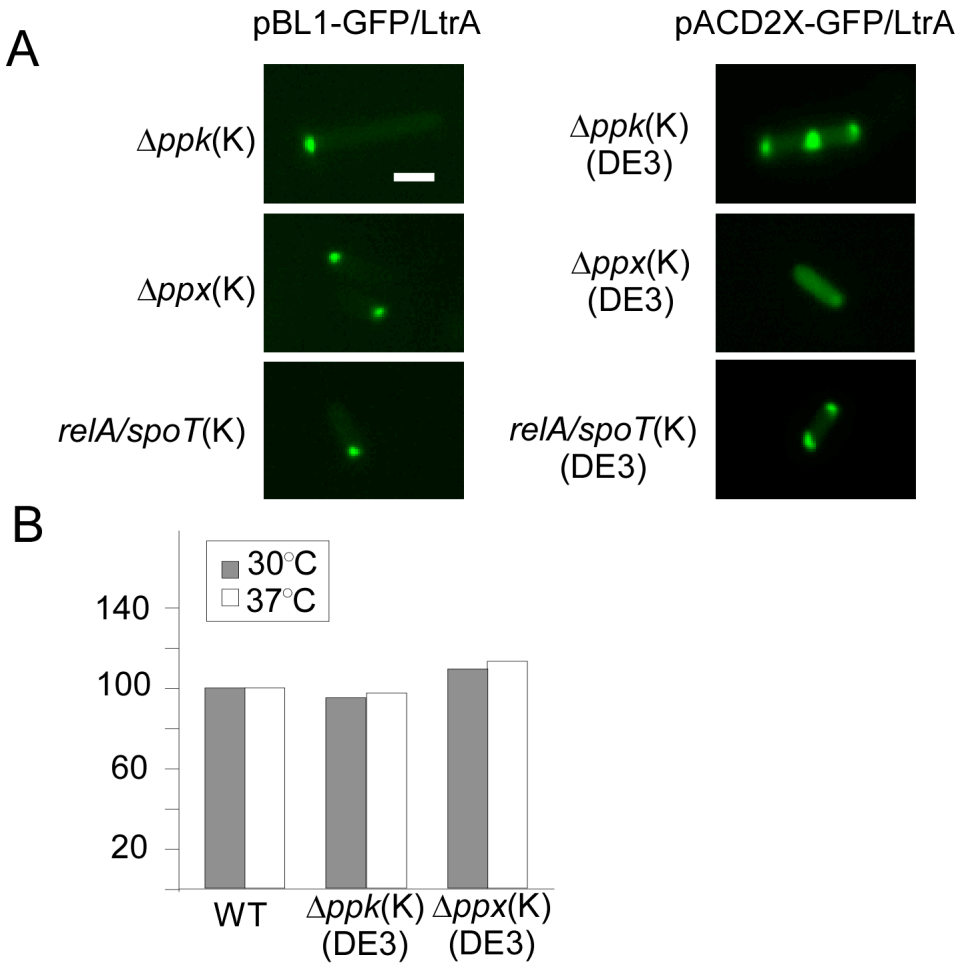

Figure 3.12: GFP/LtrA Localization and intron mobility in Keio deletions or mutant strain obtained from Genobase.

(A) Fluorescence microscopy of GFP/LtrA expressed under the Pm promoter from pBL1-GFP/LtrA (left), or under the T7 promoter from pACD2X-GFP/LtrA (right), in Keio strains (Δ*ppk*(K) and Δ*ppx*(K)) and mutant (*relA*/*spoT*(K)) obtained from Genobase. A T7 RNA polymerase gene was introduced into the strains by λ(DE3) lysogenization. The localization patterns represent the most frequent seen in each strain. Scale bar = 2 μm. (B) Mobility of the wild-type Ll.LtrB-ΔORF intron in wild-type (WT) HMS174(DE3) and knockout strains from Keio collection (Δ*ppk*(K) and Δ*ppx*(K)) at 30°C (gray bars) or 37°C (white bars).

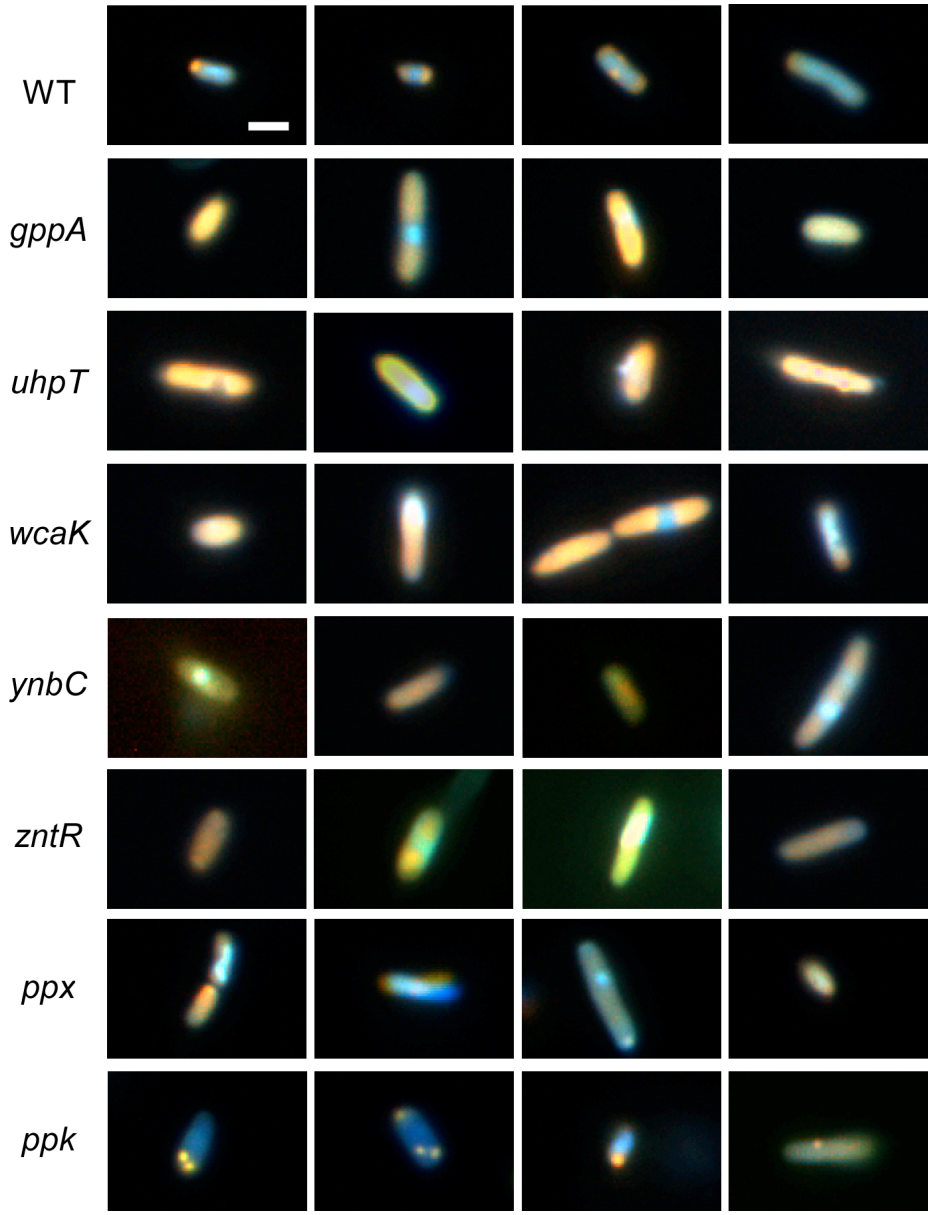

Figure 3.13: Fluorescence microscopy of poly(P) in wild-type HMS174(DE3) and disruptant strains.

DAPI staining (25 μg/ml) showing poly(P) (yellow or orange) and DNA (blue) in the wild-type (WT) HMS174(DE3), disruptants (*gppA*, *uhpT*, *wcaK*, *ynbC*, and *zntR*) identified by cell array screening, and *ppx* and *ppk* disruptants generated by targetrons in HMS174(DE3) (Ppk-1140a, Ppx-1051a). Scale bar = 2 μm.

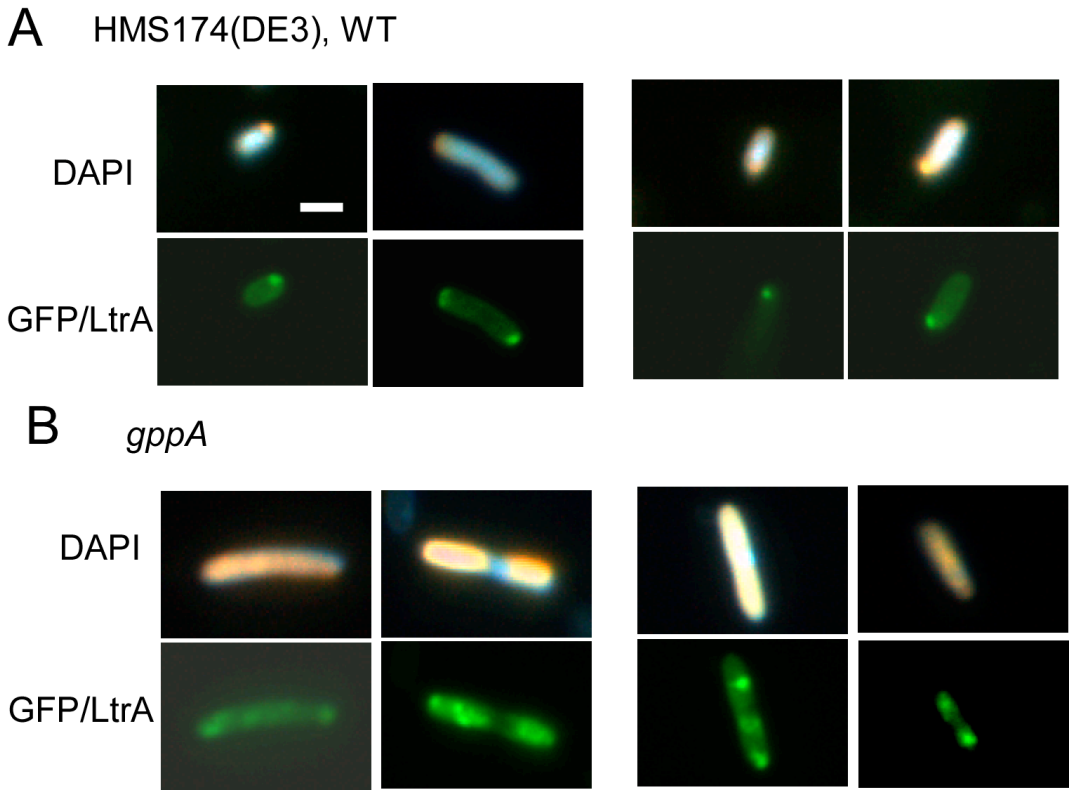

Figure 3.14: Fluorescence microscopy of GFP/LtrA and poly(P) in wild-type (WT) HMS174(DE3) and the *gppA* disruptant strain.

Wild-type (WT) (A) or the *gppA* disruptant (B) cells showing poly(P) (yellow or orange) and DNA (blue) from DAPI staining (25 μg/ml), and GFP/LtrA (green) expressed under the T7 promoter from pACD2X-GFP/LtrA. In (A, B), left, cells showing GFP/LtrA localization similar to poly(P) (80% of cells with poly(P)); right, cells showing GFP/LtrA localization different from poly(P) (20% of cells with poly(P)). Scale bar = 2 μm.

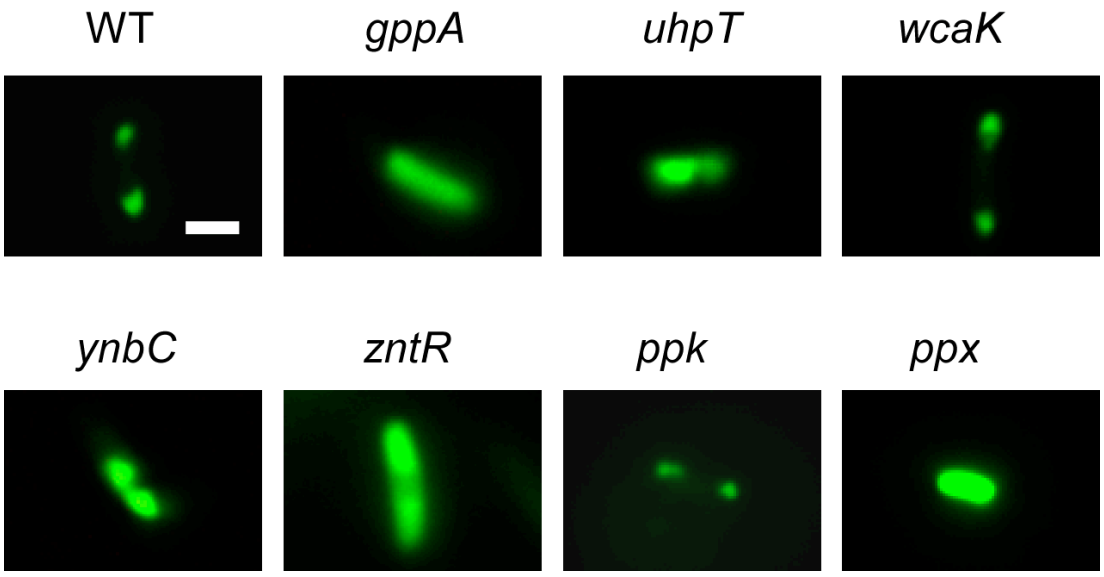

Figure 3.15: Fluorescence microscopy showing GFP/CYT-18 localization in wild-type HMS174(DE3) (WT) and the indicated disruptant strains.

Cells containing pACD2X-GFP/CYT-18 were induced with 500 μM IPTG at 30°C overnight in wild-type HMS174(DE3) (WT) cells, disruptant strains from cell array screening (*gppA, uhpT, wcaK, ynbC*, and *zntR*), and disruptant strains generated by targetron (*ppk*, *ppx*). Scale bar = 2 μm.

Chapter 4: Materials and methods

4.1 BACTERIAL STRANS AND GROWTTH CONDITIONS

Intron mobility, targeting and fluorescence microscopy experiments were carried out in *E. coli* strains HMS174(DE3) (F^- *recA hsdR*($r_{K12}^-m_{K12}^-$) Rif^R λDE3) (Novagen, San Diego, CA) and BL21(DE3) (F^- *ompT hsdS*($r_B^-m_B^-$) *gal dcm* λDE3) (Stratagene, La Jolla, CA). DH5α, used for cloning. S17-1λ*pir* (Tp^R Sm^R *recA thi pro hsdR* M^+RP4-2-Tc::Mu-Km::Tn7 λ*pir*; Simon *et al.,* 1983, obtained from Dr. Andy Ellington, Univ. Texas at Austin), used for *mariner* transposon mutagenesis; and AQ10033 (*oriC*$^+$) and AQ10060 (*oriC*$^-$) (Bates *et al.,* 1995; obtained from Dr. James Walker, Univ. Texas at Austin). Disruption/deletion strains obtained from Genobase (ecoli.naist.jp) were: *gppA* (JD24693, parent strain KP7600, F^- *lacI*^{q} *lacZ*ΔM15 *galK2 galT22* λ- *IN*(*rrnD-rrnE*)*1*), *uhpT* (JW3641), *wcaK* (JW2030), *ynbC* (JW1407), *zntR* (JW2354), *galU* (JW1224), *rfe* (JW3758), *ppk* (JW2486), *ppx* (JW2487) (from parent strain BW25113 *lacI*^{q} *rrnB*$_{T14}$ Δ*lacZ*$_{WJ16}$ *hsdR514* Δ*araBAD*$_{AH33}$ Δ*rhaBAD*$_{LD78}$), *relA*/*spoT* (AQ4319, *thr-1 leu-6 thi-1 lacY1 galK2 ara-14 xyl-5 mtl-1 proA2 his-4 argE3 rpsL-31 tsx-33 sup-37 thyA nfo-1::kan relA1 spoT thiA*). Strains from Genobase, AQ10033, and AQ10060 were lysogenized with λDE3 lysogenization kit (Novagen) to introduce an IPTG-inducible phage T7 RNA polymerase.

Standard growth conditions were LB medium at 37°C. AQ10033(DE3) and AQ10060(DE3) were grown in minimal medium (Howard-Flanders *et al.,* 1964) supplemented with 0.5% casamino acids, 1% glucose, 5 μg/ml thiamine-HCl, and 50 μg/ml each of asparagine, thymine, and tryptophan. Strain AQ4319 grows in M9 salts

glucose medium supplemented with 100 μg/ml each of asparagine, histidine, leucine, proline, threonine, tryptophan, and 2 μg/ml thiamine, 20 μg/ml thymine. Antibiotics were used at the following concentrations: ampicillin, 100 μg/ml; chloramphenicol, 25 μg/ml; kanamycin, 40 μg/ml; rifampicin, 100 μg/ml; spectinomycin, 100 μg/ml; tetracycline, 25 μg/ml, trimethoprim, 10 μg/ml.

For intron induction, overnight cultures were diluted 1:100 into fresh medium, grown 2-3 h at 37°C to $O.D._{600}$ = 0.2-0.3, and induced with IPTG (isopropyl-beta-D-thiogalactoside; GBT, St. Louis, MO), 0.2% arabinose (Sigma, St. Louis, MO), or 2 μM *m*-toluic acid (Sigma), under conditions indicated for individual experiments.

Lactococcus lactis strain NZ9800 and its *ΔltrB* derivative were grown overnight without aeration in M17 + 1% glucose (M17G) medium at 30°C (plus 10 μg/ml tetracycline for NZ9800*ΔltrB*). For cells transformed with pLE-RIG-GFP/LtrA, chloramphenicol was added at 10 μg/ml. For induction of plasmid expression, overnight cultures were diluted 1:100 and grown at 30°C until $O.D._{600}$ = 0.3, then 1 ml of cells was induced with 25 ng/ml nisin (Sigma) for 3 h at 30°C.

4.2 RECOMBINANT PLASMIDS

The intron-donor plasmid pACD2X contains a 0.9-kb Ll.LtrB-ΔORF intron and short flanking exons, cloned downstream of a T7*lac* promoter, in a Cam^R vector pACYC184; the intron contains an additional T7 promoter inserted in DIV, and the LtrA protein is expressed from a position just downstream of the 3' exon (Karberg *et al.,* 2001). The recipient plasmid pBRR3-ltrB contains a 45-bp Ll.LtrB intron target site (position -30 to +15 from the intron-insertion site) cloned upstream of a promoterless tet^R gene in a pBR322-based vector carrying an amp^R gene (Guo *et al.,* 2000; Karberg *et al.,* 2001).

pACD2X-GFP/LtrA and pACD2X-LtrA/GFP are derivatives of pACD2X in which GFPuv (an enhanced GFP variant, Crameri *et al.,* 1996), is fused in-frame to the N- or C-terminus of the LtrA ORF, respectively. pACD2X-GFP/LtrA was constructed by combining three PCR products: the GFPuv ORF amplified from pGFPuv (Clontech, Palo Alto, CA) with primers 5'NGFPF (primer sequences are in Table 4.1) and 3'NGFPR; the 3' end of the Ll.LtrB-ΔORF intron and downstream region amplified from pACD2X with primers 105 and 5'NLtrBR; and LtrA ORF nts 4-624 amplified from pACD2X with primers 3'NLtrAF and LtrABspMI. The three PCR products were amplified together with primers 105 and LtrABspMI, yielding a 2-kb product, which was digested with *Pst*I + *Bsp*MI (all the restriction enzymes are from New England Biolabs, Beverly, MA) and swapped for the corresponding fragment of pACD2X.

pACD2X-LtrA/GFP was constructed by combining the GFPuv ORF amplified from pGFPuv with primers 5'CGFPF and 3'CGFPR, and LtrA ORF nts 1460-1796 amplified from pACD2X with primers P8A and 5'CLtrAR. The two PCR products were amplified together with primers P8A and 3'CGFPR, yielding a 1.1-kb product, which was digested with *Aat*II + *Stu*I and swapped for the *Aat*II/*Eco*RV fragment of pACD2X.

pACSD2-GFP/LtrA contains the N-terminal GFP/LtrA fusion from plasmid pACD2X-LtrA/GFP in place of the LtrA ORF in intron-donor plasmid pACSD2, which carries a spc^R instead of a cam^R marker (Smith *et al.,* 2005).

pAC-GFP/LtrA is a derivative of pACD2X-GFP/LtrA which expresses the LtrA protein without the Ll.LtrB-ΔORF intron, and was constructed by digesting pACD2X-GFP/LtrA with *Xba*I + *Pst*I, filling-in with phage T4 DNA polymerase (New England Biolabs), and blunt-end ligation with T4 DNA ligase (New England Biolabs).

pAC-GFP is a matched construct, which expresses GFPuv under the T7 promoter. It was constructed by PCR amplifying the GFP ORF from pGFPuv with primers

5'NGFPF and 3'CGFPR, then digesting the PCR product with *Brs*GI + *Stu*I, and swapping for the 2.2-kb *Bsr*GI/*Eco*RV fragment of pAC-GFP/LtrA.

pAC-GFP/LtrA(2-200), pAC-GFP/LtrA(201-400), pAC-GFP/RT(70-361), and pAC-GFP/En(543-599) express N-terminal GFP fusions with the indicated LtrA subsegments. pAC-GFP/LtrA(2-200), pAC-GFP/LtrA(201-400), pAC-GFP/RT(70-361), and pAC-GFP/En(543-599) were derived from pACD2X-GFP/LtrA by PCR with primers that link the C-terminus of GFPuv to the indicated N-terminal amino acid of the LtrA segment and add a TGA stop codon at the end. The segment containing the Ll.LtrB intron was then deleted from each construct, as described above for pAC-GFP/LtrA.

pLE-LtrA/RIG contains the Ll.LtrB intron and short flanking exons cloned downstream of an inducible *nisA* promoter in the *E. coli*/*L. lactis* shuttle vector pSKH1 (Ichiyanagi *et al*., 2002). The intron carries a kan^R-RIG marker in intron domain IV for use in intron mobility assays. pLE-RIG-GFP/LtrA is a derivative expressing an N-terminal GFP/LtrA fusion, made with GFPrft. The latter is a derivative of GFPuv with improved folding and solubility properties (pBAD-GFPrft, provided by Dr. George Georgiou, Univ. Texas at Austin), which we found performed better in *L. lactis* than did GFPuv.

pLE-RIG-GFP/LtrA was constructed by combining three PCR products: the GFPrft ORF amplified from pBAD-GFPrft with primers 5'NGFPF and RFTR1; a segment from the *Afl*II site upstream of the *nisA* promoter to Ll.LtrB nt 574 amplified from pLE-LtrA/RIG with primers LEAfl2 and 5'NLtrBR; and Ll.LtrB nts 875-1198 amplified from pLE-LtrA/RIG with primers RFTLtrAF1 and LtrABspMI. The three PCR products were amplified together with primers LEAflII and LtrABspMI, yielding a 2.8-kb product, which was digested with *Afl*II + *Bsa*I and swapped for the corresponding

fragment of pLE-LtrA/RIG. pLE-RIG-GFP is a matched plasmid that expresses GFPrft from within intron domain IV.

pBAD24-*icsA*$_{507\text{-}620}$*::gfp* (Janakiraman & Goldberg, 2004), which expresses IcsA$_{507\text{-}620}$-GFP from the pBAD (arabinose) promoter, was obtained from Dr. Marcia Goldberg (Massachusetts General Hospital). pBAD24-*icsA*$_{507\text{-}620}$*::gfpt* expresses IcsA$_{507\text{-}620}$ linked to a truncated, non-fluorescent GFP. pBAD24-*icsA*$_{507\text{-}620}$*::gfpt* was constructed via PCR of pBAD24-*icsA*$_{507\text{-}620}$*::gfp* with primer 5'NGFPF and GFP154PstI. The PCR product was digested with *Bsr*GI + *Pst*I and swapped for the corresponding fragment of pBAD24-*icsA*$_{507\text{-}620}$*::gfp*.

pACD2X-IcsA and pACD2X-IcsAΔSP are derivatives of pACD2X expressing full-length IcsA and IcsA with its signal peptide deleted, respectively. pACD2X-IcsA was constructed by PCR amplifying the IcsA ORF of *Shigella flexneri* 2457T DNA (obtained from Shelly Payne, University of Texas at Austin) with primers 5IcsAF and 3IcsAR. The 3.3-kb PCR product was digested with *Hin*dIII + *Eco*RV and swapped for the corresponding fragment of pACD2X. pACD2X-IcsAΔSP was made similarly with primer 5IcsAFΔSP instead of 5IcsAF.

pSC189, used for *mariner* transposon mutagenesis, employs a hyperactive C9 transposase and *mariner* transposon carrying a *kan*^{R} marker for selection of integration events (Chiang & Rubin, 2002). The plasmid contains an *ori*R6K replication origin which requires the *pir* gene to replicate (here we use *E. coli* strain S17-1λ*pir*) and is conjugated from that strain into *E. coli* HMS174(DE3), where it is unable to replicate and thus rapidly lost after expressing the transposon cassette.

Retargeted Ll.LtrB-ΔORF introns (targetrons) used to disrupt the *himA*, *galU*, *rfe*, *ppk*, and *ppx* genes were constructed in plasmid pACD-Kan^{R}-RAM, with the intron carrying a Kan^{R}-RAM marker to facilitate the detection of integration events (Perutka *et*

al., 2004, Zhao & Lambowitz, 2005). Targeting of the intron to insert at sites within these genes was done with the aid of a computer algorithm, which identifies potential Ll.LtrB insertion sites and designs PCR primers to modify the intron's EBS2, EBS1, and δ sequences to base pairing optimally to those sites; the IBS2 and IBS1 sequences in the donor plasmid are also modified to be complementary to the retargeted EBS2 and EBS1 sequences for efficient RNA splicing (Perutka *et al.,* 2004). The required modifications were introduced into the intron donor plasmid by a two-step PCR, as described (Guo *et al.,* 2000, Karberg *et al.,* 2001), using the following pairs of primers: himA140aIBS + ASEBS2; himA140aEBS1 + himA140aEBS2; galU257aIBS + ASEBS2; galU257aEBS1 + galU257aEBS2; rfe252sIBS + ASEBS2; rfe252sEBS1 + rfe252sEBS2; ppk1140aEBS1 + ppk1140aEBS2; ppk1140aIBS + ASEBS2; ppx22aEBS1 + ppx22aEBS2; ppx22aIBS + ASEBS2. In each case, the final 330-bp product with *Hind*III + *BsrG*I at the ends was digested with *Hind*III + *BsrG*I and swapped for the corresponding fragment of the donor plasmid.

pBL1-GFP/LtrA expresses GFP/LtrA under the Pm promoter and was constructed by inserting GFP/LtrA *Hind*III/*EcoR*V fragment of pACD2X-GFP/LtrA into vector pBL1 (Yao & Lambowtiz, 2007) digested with *Hind*III + *Xho*I and blunt-ended with T4 DNA polymerase (New England Biolabs).

In all constructs, regions subjected to PCR were sequenced to insure that no adventitious mutations had been introduced.

4.3 INTRON MOBILITY ASSAYS

Intron mobility was assayed using an *E. coli* two-plasmid system (Figure 1.4A, Guo *et al.,* 2000, Karberg *et al.,* 2001). The Cam^R intron-donor plasmid pACD2X and Amp^R recipient plasmid pBRR3-ltrB were co-transformed into *E. coli* HMS174 (DE3),

and cells were grown overnight at 37°C in LB medium containing chloramphenicol and ampicillin. The overnight culture was then diluted 1:100 into fresh LB with the same antibiotics, grown 2-3 h at 37°C to $O.D._{600}$ = 0.2-0.3, and induced with 100 μM IPTG for 1 h at 37°C or 500 μM 3h at 30°C, unless described elsewhere. After induction, cells were washed, resuspended in fresh medium, and plated at different dilutions on LB containing tetracycline plus ampicillin or ampicillin alone. Colonies were counted after overnight incubation at 37°C, and mobility frequencies were calculated as the ratio of $(Tet^R + Amp^R)/Amp^R$ colonies.

4.4 *E. COLI* CHROMOSOMAL GENES DISRUPTION BY RETARGETE GROP II INTRON

Constructs were based on pACD3-Kan^R-RAM with IBS and EBS sequences modified to target different genes (Ichiyanagi *et al.,* 2002; Zhao & Lambowitz, 2005; constructed by Jun Yao in Lambowitz lab). The Ll.LtrB targetrons have a Kan^R-RAM inserted in intron domain IV to select disruption events. The transformants were grown in LB medium with chloramphenicol overnight at 37°C, diluted 1:100 into fresh medium and grown at 37°C to $O.D._{600}$ = 0.3, and induced with 500 μM IPTG overnight at 30°C. After induction, cells were inoculated 1:25 into LB medium with kanamycin, and incubate overnight at 30°C for selection. The saturated culture after selection was plated onto LB plus kanamycin and disruptants were identified by colony PCR and confirmed by sequencing the intron-insertion junctions. After sequence confirmation, a culture was grown up in LB medium with kanamycin at 30°C overnight from single colony and stored at -80°C.

4.5 FLUORESCENCE MICROSCOPY AND CONFOCAL FLUORESCENCE MICROSCOPY

Cells from 1 ml of culture were washed and resuspended in 0.1-ml of fresh growth medium without IPTG or antibiotics, and ~1 μl was placed on a glass slide and

examined by fluorescence microscopy using a Zeiss Axioplan2 with 63X/1.40 oil DIC lens and a GFP filter. Photographs were taken with a Hamamatsu c4742-95 digital camera and processed using Photoshop. DNA was stained with DAPI (4'-6-diamidino-2-phenylindole, 10 μg/ml; Molecular Probes; Eugene, OR) for 15 min to 1 h before the end of the induction period, and cell membranes were stained with FM4-64 (2 μg/ml; Molecular Probes) for 15 min before the end of the induction period. *L. lactis* cells were chilled at 4°C for 5-6 h prior to microscopy to enhance fluorescence (Fernandez *et al.,* 2000). The fluorescence of poly(P) was observed in cells stained with 25 μg/ml of DAPI added to the culture 30 min before the end of induction. The strains obtained by transposon mutagenesis, or from the Keio collection were examined manually by fluorescence microscopy using a Leica DMIRBE microscope with 100x oil (PL APO 1.4-0.7 NA) lens and a GFP filter. Photographs were taken with a Leica DFC350 FX fluorescence camera, or Leica DFC320 FX fluorescence camera for color pictures of poly(P) fluorescence. The *mreB* mutant expressing GFP/LtrA under IPTG induction was examined by using a Leica SP2 AOBS confocal microscope. All the pictures were processed using IrfanView and Photoshop CS.

4.6 IMMUNOFLUORESCENCE MICROSCOPY

Cells from 10 ml of culture were fixed directly in growth medium by adding paraformaldehyde, glutaraldehyde, and $NaPO_4$ buffer, pH 7.4, to final concentrations of 2.4% (v/v), 0.04% (v/v), and 30 mM, respectively, and then incubating at room temperature for 10 min, and on ice for 50 min. The cells were washed three times by centrifugation in phosphate-buffered saline (PBS) at room temperature, resuspended in 50 mM glucose, 10 mM EDTA, 20 mM Tris-HCl, pH 7.5, and incubated with freshly prepared lysozyme (5 μg/ml; Sigma) for 30 min at room temperature. After washing

twice with PBS, the cells were air-dried, rehydrated with PBS, incubated for 4 min at room temperature, recentrifuged, incubated in blocking solution [2% (w/v) bovine serum albumin in PBS] for 30 min at room temperature, and recentrifuged. They were then incubated overnight at 4°C in blocking solution containing a 1:100 dilution of rabbit anti-LtrA polyclonal antibodies preparation (obtained from Dr. Gary Dunny, University of Minnesota, and pre-adsorbed to fixed untransformed HMS174(DE3)), washed 10 times with blocking solution, and incubated in the dark for 2 h at room temperature in blocking solution with a 1:1000 dilution of secondary antibody (anti-rabbit IgG-FITC; Jackson Lab, Bar Harbor, Maine). Cells were examined microscopically after ten washes with PBS.

4.7 SDS-PAGE AND IMMUNOBLOTTING

SDS-PAGE and immunoblotting were as described (Cui *et al.,* 2004). Samples containing protein from 0.05 $O.D._{600}$ units of cells were analyzed in either 7.5% polyacrylamide/0.1% SDS (GFP/LtrA) or 10% polyacrylamide/0.1% SDS (IcsA/GFP) gels. Samples were then transferred to a nitrocellulose membrane (BioRAD, Hercules, CA) via a semi-dry transfer unit (Hoefer Semiphor TE 70, Amersham Biosciences Corp., San Francisco, CA). Immunoblots were probed with a 1:1000 dilution of anti-LtrA antibodies (see above) or anti-GFP antibody (BD Biosciences; Franklin Lakes, NJ), followed by a 1:10,000 dilution of HRP goat anti-rabbit secondary antibody for LtrA antibody (Pierce, Rockford, IL) or 1:10,000 dilution of HRP goat anti-mouse secondary antibody (BioRAD). Blots were developed with SuperSignal West Pico Chemilluminescent substrate (Pierce) or Amersham ECL western blotting detection reagents (GE healthcare, Piscataway, NJ). Equal loading was confirmed by Coomassie blue staining of a parallel gel.

4.8 CONSTRUCTION OF A *MARINER* TRANSPOSON KNOCKOUT LIBRARY

E. coli strain S17-1λ*pir* carrying the *mariner* transposon delivery vector pSC189 and HMS174(DE3) were grown separately in 10 ml of LB with or without ampicillin at 37°C to $O.D._{600}$ = 0.5, then mixed, incubated at room temperature overnight without shaking, washed twice with LB, then resuspended in LB containing rifampicin and kanamycin, and grown overnight at 37°C to select for HMS174(DE3) cells containing *mariner* transposon insertions. The integration efficiency of *mariner* transposon was ~ 10^{-4}. Loss of pSC189, which carries an amp^R marker and does not replicate in HMS174(DE3), was confirmed by plating the cells on LB plates with and without ampicillin (<1% Amp^R). The knockout culture was grown in LB with kanamycin, made electroporation competent (Dower, 1990), electroporated with pACD2X-GFP/LtrA and plated on LB plates containing chloramphenicol and kanamycin. About 9,600 individual colonies were then picked into LB medium with chloramphenicol and kanamycin in 100 96-well plates, grown overnight at 37°C, and stored at -80°C. The complexity of the library was confirmed by TAIL PCR of ~60 individual isolates, which showed >92% with different *mariner* insertion sites both before and after transformation with pACD2X-GFP/LtrA (55 of 58 and 56 of 61, respectively).

4.9 CELL ARRAY CONSTRUCTION AND IMAGING

For printing cell arrays, the knockout library and a single well of wild-type HMS174(DE3) containing pACD2X-GFP/LtrA were inoculated in new 96-well plates containing fresh LB plus chloramphenicol and incubated overnight at 37°C. The cultures were then inoculated 1:10 into fresh LB medium in new 96-well plates and grown for 3 h at 37°C. In one experiment (fifty one 96-well plates), the cells were induced with 100 μM IPTG at 37°C, and in a second experiment (forty nine 96-well plates), the cells were induced 500 μM IPTG at 30°C. Culture transfer and media additions were done by a

Biomek FX Laboratory Automation Workstation (Beckman Coulter, Fullerton, CA). Cell arrays of ~5,000 knockouts and one wild-type control were printed onto microscope slides by using a custom-built DNA microarray printing arrayer. 30 copies were made for each slide, and the best two were used for analysis. After printing, the cell arrays were washed with 25% glycerol containing DAPI (15 μg/ml) and used immediately for imaging. Cell images were collected by automated microscopy, using a Nikon E800 microscope with computer-controlled *X-Y* stage and piezoelectric-positioned objective, by scanning the position of each spot, autofocusing, and capturing the image with a Coolsnap CCD camera (Photometrics, Tucson, USA). Images were stored in a custom cell microarray image database (Cellma, cellma.icmb.utexas.edu) and examined individually to identify strains with altered GFP/LtrA localization patterns.

4.10 TAIL PCR (THERMAL ASYMMETRIC INTERLACED PCR)

TAIL PCR was done on colonies re-suspended in PCR pre-mix, as described (Liu & Whittier, 1995). Integration junctions were amplified by using two nested specific primers and one degenerate primer. For *mariner* transposon insertions, the specific primer TailP1 and degenerate primer AD2 (Liu & Whittier, 1995) were used for the first PCR, and the specific primer TailP2 and degenerate primer AD2 were used for the second PCR. The final PCR product was gel-purified and sequenced using the TailP2 primer. For Ll.LtrB insertions, the specific primers used in the first and second PCR were Ell1 and Ell2, respectively, and the degenerate primer was again AD2. The final TAIL PCR products were sequenced by using primer Ell3.

4.11 SOUTHERN HYBRIDIZATION AND INSERTION JUNCTION PCR FOR *MARINER* TRANSPOSON-INSERTIONS IN DISRUPTANTS

Genomic DNA from each disruptant and as well as wild type HMS174(DE3) cells was purified by using a genomic DNA Isolation Kit (Qiagen, Valencia, CA). 10 μg of genomic DNA was digested by *Xcm*I, *Xma*I, and *Sac*II (New England Lab, 60 U each) at 37°C overnight and loaded onto an 8% agarose gel. After running, the gel was blotted to a nylon transfer membrane (Magna, 0.45 μm; Osmonics Inc., Minnetonka, MN) and hybridized with a ^{32}P-labeled probe corresponding to *mariner* transposon positions 1385-1868. The probe was generated by PCR with primers Mar3200 and Mar-3650, followed by labeling with [α-^{32}P] dTTP (3000 Ci/mmol; NEN Life Science Products, Boston, MA), using a High Prime DNA labeling kit (Roche, Indianapolis, IN). The blots were scanned with Typhoon Trio fluorescence scanner (GE healthcare).

The 5'- and 3' -junctions of *mariner* transposon insertions in the disruptants were also amplified by PCR with transposon primers (TAIL P2, Mar-2020) and combined with specific primers (*gppA*, gppA-3end, gppA850; *uhpT*, ade1550, uhpT1270; *wcaK*, wcaK-300, wcaK50; *ynbC*, ynbC990, ynbC-1670; *zntR*, zntR-420, yhdN30) respectively.

4.12 CHROMOSOMAL DISTRIBUTION OF LL.LTRB INSERTION SITES IN WILD-TYPE AND DISRUPTANT STRAINS

Cells were transformed with a pACD3-RAM library of Ll.LtrB-ΔORF introns with randomized EBS2, EBS1, and δ sequences (Zhong *et al.,* 2003). The Ll.LtrB introns in the library have a Tp^{R}-RAM inserted in intron domain IV to select insertion events. The transformants were grown in LB medium containing chloramphenicol and kanamycin overnight at 37°C, inoculated 1:100 into fresh medium, grown at 37°C to O.D.$_{600}$ = 0.3, then induced with 500 μM IPTG overnight at 30°C. After induction, cells were washed by centrifugation and resuspention in fresh Mueller-Hinton medium (MH

medium, BD Biosciences), then plated onto MH medium with trimethoprim and thymine (Zhong *et al.,* 2003) and incubated at 30°C for 48 h. Ll.LtrB intron insertion sites were identified by TAIL PCR of DNA from individual colonies (Liu & Whittier, 1995) and sequenced as described.

Table 4.1: List of oligonucleotides.

Name	Sequence
105	5’- TGAAACCAACAATGGCAATTTTAG
3’CGFPR	5’- CGGCCGACTAGTAGGCCTATTATTTTTGACACC
3’NGFPR	5’-CTGATTCTTTCTAAAATTGCTTTGTAGAGCTCATCCATGGC
3’NLtrAF	5’- ACACATGGCATGGATGAGCTCTACAAAGCAATTTTAGAAAGAATCAGTAAAAATTC
3IcsAR	5’- CCCCTCGAGTCAGAAGGTATATTTCACACCC
5’CGFPF	5’- CGTGATTCATAAACACAAGAGTAAAGGAGAAGAACTTTTCAC
5’CLtrAR	5’- CTTCTCCTTTACTCTTGTGTTTATGAATCACGTGACG
5’NGFPF	5’- TTGATTAGGGAGGAAAACCTCAAAATGAGTAAAGGAGAAGAACTTTTCAC
5’NLtrBR	5’- CCTTTACTCATTTTGAGGTTTTCCTCCCTAATC
5IcsAF	5’- GGGAGGAAAACCTCAAACAACCACTTACTGATAATATAGTGCATG
5IcsAFΔSP	5’- GAGGAAAACCTCAAAATGACTCCTCTTTCGGGTACTCAAGAACTTC
AD2	5’- NGTCGASWGANAWGA
ade1550	5’- TGCCGTTACCCATTGCCGGGCTGATGAGC
ASEBS2	5’- AACCGAAATTAGAAACTTGCGTTCAG
Ell1	5’- CTGATTAACATTGCGACTCAGTCGTACCC
Ell2	5’- CAACCGTGCTCTGTTCCCGTATCAG
Ell3	5’- GGTTGGCTGTTTTCTGTGTTATCTTACAGAG
galU257aEBS1	5’- CAGATTGTACAAATGTGGTGATAACAGATAAGTCGTTTTACATAACTTACCTTTCTTTGT
galU257aEBS2	5’- TGAACGCAAGTTTCTAATTTCGATTTCAACTCGATAGAGGAAAGTGTCT
galU257aIBS	5’- AAAAAAGCTTATAATTATCCTTAGTTGACGTTTTAGTGCGCCCAGATAGGGTG
gppA-3end	5’- GCGTCAGCATCGCATCCGGCAC
gppA850	5’- CAGTGTATGACCCTGGCGGGCGG
GFP154PstI	5’- CGTTCTGCAGTCATGCCGTGATGTATACATTGTGTG
HimA140aEBS1	5’- CAGATTGTACAAATGTGGTGATAACAGATAAGTCAACCAGAGTAACTTACCTTTCTTTGT
HimA140aEBS2	5’- TGAACGCAAGTTTCTAATTTCGGTTTGGTATCGATAGAGGAAAGTGTCT
HimA140aIBS	5’- AAAAAAGCTTATAATTATCCTTATACCACAACCAGGTGCGCCCAGATAGGGTG
LEAfl2	5’- CTGGCAGCTTAAGTTTGCTTCTAAGTCTTATTTCC
LtrABspMI	5’- CCGCTGTAAGTTTTGTGATACTGCCAG
Mar-2020	5’- CCCTTATAAATCAAAAGAATAGACCGAGATAGG
Mar3200	5’- GGGTGGAGAGGCTATTCGGCTATGACTGGGC
Mar-3650	5’- CCTTGAGCCTGGCGAACAGTTCGGCTGG

Table 4.1: List of oligonucleotides (continued).

Name	Sequence
P8A	5’- TGGGGGATCCCGTATGAGATAAAGC
ppk1140aEBS1	5’- CAGATTGTACAAATGTGGTGATAACAGATAAGTCGATGCGTGTAACTTACCTTTCTTTGT
ppk1140aEBS2	5’- TGAACGCAAGTTTCTAATTTCGATTTCGACTCGATAGAGGAAAGTGTCT
ppk1140aIBS	5’- AAAAAAGCTTATAATTATCCTAGTCGACGATGCGGTGCGCCCAGATAGGGTG
ppx1051aEBS1	5’- CAGATTGTACAAATGTGGTGATAACAGATAAGTCTCGGTTGCTAACTTACCTTTCTTTGT
ppx1051aEBS2	5’- TGAACGCAAGTTTCTAATTTCGGTTGCTGGTCGATAGAGGAAAGTGTCT
ppx1051aIBS	5’- AAAAAAGCTTATAATTATCCTTACCAGCCTCGGTTGTGCGCCCAGATAGGGTG
rfe252sEBS1	5’- CAGATTGTACAAATGTGGTGATAACAGATAAGTCCTTGTTTTTAACTTACCTTTCTTTGT
rfe252sEBS2	5’- TGAACGCAAGTTTCTAATTTCGATTACACCTCGATAGAGGAAAGTGTCT
rfe252sIBS	5’- AAAAAAGCTTATAATTATCCTTAGGTGTCCTTGTTGTGCGCCCAGATAGGGTG
RFTLtrAF1	5’- GGCATGGATGAACTATACAAAGCAATTTTAGAAAGAATCAGTAAAAATTC
RFTR1	5’- GATTCTTTCTAAAATTGCTTTGTATAGTTCATCCATGCCATGTGTAATCCC
TailP1	5’- GTTCTTCTGAGCGGGACTCTGGGG
TailP2	5’- CGGCCGCGAAGTTCCTATTCCG
uhpT1270	5’- GGCTGGGCAGGCACCTTCGCCGCGC
wcaK50	5’- GGGCAACCACACTTGCGGCAATCG
wcaK-300	5’- GTCGGTGAATCCCTGGGCGATGGCG
yhdN30	5’- GGCAGAGCGCCATATAGCAGAAGCGC
ynbC990	5’- TTGTTCGGGACGCACTCTTCGGGCCTGC
ynbC-1670	5’- TCACGCACGAGTGAATCCATCTCCCC
zntR-420	5’- CCACTCTTAACGCCACTCGCCCCTTGTTC

Bibliography

Aizawa, Y., Xiang, Q., Lambowitz, A.M. & Pyle, A.M. (2003) The pathway for DNA recognition and RNA integration by a group II intron retrotransposon. *Mol. Cell.* **11,** 795-805.

Akiyama, M., Crooke, E. & Kornberg, A. (1992) The polyphosphate kinase gene of *Escherichia coli*. *J. Biol. Chem.* **267,** 22556-22561.

Akiyama, M., Crooke, E. & Kornberg, A. (1993) An exopolyphosphatase of *Escherichia coli*. *J. Biol. Chem.* **268,** 633-639.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L. & Mori, H. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* Published online, doi: 10.1038/msb4100050.

Bates, D.B., Asai, T., Cao, Y., Chambers, M.W., Cadwell, G.W., Boye, E. & Kogoma, T. (1995) The DnaA box R4 in the minimal *oriC* is dispensable for initiation of *Escherichia coli* chromosome replication. *Nucl. Acids. Res.* **23,** 3119-3125.

Belfort, M., Derbyshire, V., Parker, M.M., Cousineau, B. & Lambowitz, A.M. (2002) Mobile introns: pathways and proteins. In *Mobile DNA II*, eds. Craig, N.L. *et al.* (ASM Press Publishers, Washington D.C.), pp. 761-783.

Beauregard, A., Chalamcharla, V.R., Piazza, C.L., Belfort, M. & Coros, C.J. (2006) Bipolar localization of the group II intron Ll.LtrB is maintained in *Escherichia coli* deficient in nucleoid condensation, chromosome partitioning and DNA replication. *Mol Microbiol.* **62,** 709-722.

Blattner, F.R., Plunkett, G. 3rd, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. & Shao, Y. (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* **277,** 1453-1474.

Blocker, F.J., Mohr, G., Conlan, L.H., Qi, L., Belfort, M. & Lambowitz, A.M. (2005) Domain structure and three-dimensional model of a group II intron-encoded reverse transcriptase. *RNA* **11,** 14-28.

Brandon, L.D., Goehring, N., Janakiraman, A., Yan, A.W., Wu, T., Beckwith, J. & Goldberg, M.B. (2003) IcsA, a polarly localized autotransporter with an atypical

signal peptide, uses the Sec apparatus for secretion, although the Sec apparatus is circumferentially distributed. *Mol. Microbiol.,* **50,** 45-60.

Brown, M.R. & Kornberg, A. (2004) Inorganic polyphosphate in the origin and survival of species. *Proc. Natl. Acad. Sci. USA*. **101,** 16085-16087.

Carrió, M.M. & Villaverde, A. (2005) Localization of chaperones DnaK and GroEL in bacterial inclusion bodies. *J. Bacteriol.* **187,** 3599-3601.

Charles, M., Perez, M., Kobil, J.H. & Goldberg, M.B. (2001) Polar targeting of *Shigella* virulence factor IcsA in Enterobacteriacae and *Vibrio*. *Proc. Natl. Acad. Sci. USA* **98,** 9871-9876.

Chiang, S.L. & Rubin, E.J. (2002) Construction of a *mariner*-based transposon for epitope-tagging and genomic targeting. *Gene* **296,** 179-185.

Coppola, G. & Geschwind, D.H. (2006) Microarrays and the microscope: balancing throughput with resolution. *J. Physiol.* **575,** 353-359.

Coros, C.J., Landthaler, M., Piazza, C.L., Beauregard, A., Esposito, D., Perutka, J., Lambowitz, A.M. & Belfort, M. (2005) Retrotransposition strategies of the *Lactococcus lactis* Ll.LtrB group II intron are dictated by host identity and cellular environment. *Mol. Microbiol.* **56,** 509-524.

Crameri, A., Whitehorn, E.A., Tate, E. & Stemmer, W.P. (1996) Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nature Biotech.* **14,** 315-319.

Cui, X., Matsuura, M., Wang, Q., Ma, H. & Lambowitz, A.M. (2004) A group II intron-encoded maturase functions preferentially i*n cis* and requires both the reverse transcriptase and X domains to promote RNA splicing. *J. Mol. Biol.* **340,** 211-231.

Dower, W.J. "Electroporation of bacteria: a general approach to genetic transformation," in *Genetic Engineering –Principles and Methods*, 1990, vol. 12, pp. 275-296, Plenum Publishing Corp., New York.

Draper, G.C. & Gober, J.W. (2002) Bacterial chromosome segregation. *Annu. Rev. Microbiol.* **56,** 567-597.

Eickbush, T. H. (1994) Origin and evolutionary relationships of retroelements. In *The Evolutionary Biology of Viruses*, Morse, S. S., Ed. (New York: Raven Press, Ltd.), pp.121-157.

Eskes, R., Yang, J., Lambowitz, A. M. & Perlman, P. S. (1997) Mobility of yeast mitochondrial group II introns: engineering a new site specificity and retrohoming via full reverse splicing. *Cell* **88,** 865-874.

Fedorova, O., Mitros, T. & Pyle, A.M. (2003) Domains 2 and 3 interact to form critical elements of the group II intron active site. *J.Mol.Biol.* **330,** 197-209.

Fernandez de Palencia, P., Nieto, C., Acebo, P., Espinosa, M. & Lopez, P. (2000) Expression of green fluorescent protein in *Lactococcus lactis*. *FEMS Microbiol. Lett.* **183,** 229-234.

Gentry, D.R. & Cashel, M. (1996) Mutational analysis of the *Escherichia coli spoT* gene identifies distinct but overlapping regions involved in ppGpp synthesis and degradation. *Mol. Microbiol.* **19,** 1373-1384.

Goldberg, M.B., Barzu, O., Parsot, C. & Sansonetti, P.J. (1993) Unipolar localization and ATPase activity of IcsA, a *Shigella flexneri* protein involved in intracellular movement. *J. Bacteriol.* **175,** 2189-2196.

Gomez-Garcia, M.R. & Kornberg, A. (2004) Formation of an actin-like filament concurrent with the enzymatic synthesis of inorganic polyphosphate. *Proc. Natl. Acad. Sci. USA*. **101,** 15876-15880.

Gordon, G.S., Shivers, R.P. & Wright, A. (2002) Polar localization of the *Escherichia coli oriC* region is independent of the site of replication initiation. *Mol. Microbiol.* **44,** 501-507.

Guo, H., Karberg, M., Long, M., Jones, J.P. 3rd, Sullenger, B. & Lambowitz, A.M. (2000) Group II introns designed to insert into therapeutically relevant DNA target sites in human cells. *Science* **289,** 452-457.

Hall, J.A. & Maloney, P.C. (2001) Transmembrane segment 11 of UhpT, the sugar phosphate carrier of *Escherichia coli*, is an alpha-helix that carries determinants of substrate selectivity. *J. Biol. Chem.* **276,** 25107-25113.

Howard-Flanders, P., Simson, E. & Theriot, L. (1964) A locus that controls filament formation and sensitivity to radiation in *Escherichia* coli K-12. *Genetics* **49,** 237-246.

Hung, C., Peccia, J., Zilles, J.L. & Noguera, D.R. (2002) Physical enrichment of polyphosphate-accumulating organisms in activated sludge. *Water Environ. Res.* **74,** 354-361.

Ichiyanagi, K., Beauregard, A., Lawrence, S., Smith, D., Cousineau, B. & Belfort, M. (2002) Retrotransposition of the Ll.LtrB group II intron proceeds predominantly via reverse splicing into DNA targets. *Mol. Microbiol.* **46,** 1259-1272.

Ichiyanagi, K., Beauregard, A. & Belfort, M. (2003) A bacterial group II intron favors retrotransposition into plasmid targets. *Proc. Natl. Acad. Sci. USA* **100,** 15742-15747.

Jain, S., Ulsen, P.V., Benz, I., Schmidt, M.A., Fernandez, R., Tommassen, J. & Goldberg, M.B. (2006) Polar localization of the autotransporter family of large bacterial virulence proteins. *J. Bacteriol.* **188,** 4841-4850.

Janakiraman, A. & Goldberg, M.B. (2004) Evidence for polar positional information independent of cell division and nucleoid occlusion. *Proc. Natl. Acad. Sci. USA* **101,** 835-840.

Karberg, M., Guo, H., Zhong, J., Coon, R., Perutka, J. & Lambowitz, A.M. (2001) Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria. *Nature Biotech.* **19,** 1162-1167.

Keasling, J.D., Bertsch, L. & Kornberg, A. (1993) Guanosine pentaphosphate phosphohydrolase of *Escherichia coli* is a long-chain exopolyphosphatase. *Proc. Natl. Acad. Sci. USA.* **90,** 7029-7033.

Kittle, J.D. Jr., Mohr, G., Gianelos, J.A., Wang, H. & Lambowitz, A.M. (1991) The *Neurospora* mitochondrial tyrosyl-tRNA synthetase is sufficient for group I intron splicing in vitro and uses the carboxy-terminal tRNA-binding domain along with other regions. *Genes Dev.* **5,** 1009-1021.

Kornberg, A. (1995) Inorganic polyphosphate: toward making a forgotten polymer unforgettable. *J. Bacteriol.* **177,** 491-496.

Kristensen, O., Laurberg, M., Liljas, A., Kastrup, J.S., & Gajhede, M. (2004) Structural characterization of the stringent response related exopolyphosphatase/guanosine pentaphosphate phosphohydrolase protein family. *Biochemistry* **43,** 8894-8900.

Kruse, T., Bork-Jensen, J. & Gerdes, K. (2005) The morphogenetic MreBCD proteins of *Escherichia coli* form an essential membrane-bound complex. *Mol. Microbiol.* **55,** 78-89.

Kuroda A., Murphy, H., Cashel, M. & Kornberg, A. (1997) Guanosine tetra- and pentaphosphate promotes accumulation of inorganic polyphosphate in *Escherichia coli*. *J. Biol. Chem.* **272,** 21240-21243.

Lambowitz, A.M. & Zimmerly, S. (2004) Mobile group II introns. *Annu. Rev. Genet.* **38,** 1-35.

Lambowitz, A.M., Mohr, G. & Zimmerly, S. (2005) In *Nucleic Acids and Molecular Biology, Vol. 16: Homing Endonucleases and Inteins,* eds Belfort, M. *et. al.* (Springer, Heidelberg), pp. 121-145.

Liu, Y.G., & Whittier, R.F. (1995) Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. *Genomics* **25,** 674-681.

Luban, C., Beutel, M., Stahl, U. & Schmidt, U. (2005) Systematic screening of nuclear encoded proteins involved in the splicing metabolism of group II introns in yeast mitochondria. *Gene* **385,** 72-79.

Lybarger, SR and Maddock, JR. (2001) Polarity in action: asymmetric protein localization in bacteria. *J. Bacteriol.* **183,** 3261-3267.

Matasuura, M., Saldanha, R., Ma, H., Wank, H., Yang, J., Georg, M., Cavanagh, S., Dunny, G.M., Belfort, M. & Lambowitz, A.M. (1997) A bacterial group II intron encoding reverse transcriptase, maturase, and DNA endonuclease activities: biochemical demonstration of maturase activity and insertion of new genetic information within the intron. *Genes Dev.* **11,** 2910-2924.

Michel, F. & Ferat, J. (1995). Structure and activity of group II introns. *Annu. Rev. Biochem.* **64,** 435-461.

Mills, D., McKay, L.L. & Dunny, G.M. (1996) Splicing of a group II intron involved in the conjugative transfer of pRS01 in *Lactococci*. *J. Bacteriol.* **178,** 3531-3538.

Mills, D.A., Manias, D.A., McKay, L.L. & Dunny, G.M. (1997) Homing of a group II intron from *Lactococcus lactis* subsp. *lactis* ML3. *J. Bacteriol.* **179,** 6107-6111.

Mohr, G., Smith, D., Belfort, M. & Lambowitz, A.M. (2000) Rules for DNA target-site recognition by a lactococcal group II intron enable retargeting of the intron to specific DNA sequences. *Genes Dev.* **14**, 559-573.

Newberry, K.J. & Brennan, R.G. (2004) The structural mechanism for transcription activation by MerR family member multidrug transporter activation, N terminus. *J. Biol. Chem.* **279,** 20356-20362.

Niki, H. & Hiraga, S. (1998) Polar localization of the replication origin and terminus in *Escherichia coli* nucleoids during chromosome partitioning. *Genes Dev.* **12,** 1036-1045.

Niki, H., Yamaichi, Y. & Hiraga, S. (2000) Dynamic organization of chromosomal DNA in *Escherichia coli*. *Genes Dev.* **14,** 212-223.

Nilsen, T, Yan, AW, Gale, G, & Goldberg, MB (2005) Presence of multiple sites containing polar material in spherical *Escherichia coli* cells that lack MreB. *J. Bacteriol.* **187**, 6187-6196.

Noah, J.W. & Lambowitz, A.M. (2003) Effects of maturase binding and Mg^{2+} concentration on group II intron RNA folding investigated by UV cross-linking. *Biochemistry* **42,** 12466-12480.

Onogi, T., Ohsumi, K., Katayama, T. & Hiraga, S. (2002) Replication-dependent recruitment of the β-subunit of DNA polymerase III from cytosolic spaces to replication forks in *Escherichia coli. J. Bacteriol.* **184,** 867-870.

Perutka, J., Wang, W., Goerlitz, D., and Lambowitz, A.M. (2004) Use of computer designed group II introns to disrupt *Escherichia coli* DexH/D-box protein and DNA helicase genes. *J. Mol. Biol.* **336,** 421-439.

Puskarova, A., Ferianc, P., Kormanec, J., Homervora, D., Farewell, A. & Nystrom, T. (2002) Regulation of *yodA* encoding a novel cadmium-induced protein in *Escherichia coli*. *Microbiology* **148,** 3801-3811.

Pyle, A.M., & Lambowitz, A.M. Group II introns: ribozymes that splice RNA and invade DNA. In: The RNA World, 3rd Edition (R.F. Gesteland, T.R. Cech, and J.F. Atkins, Editors), Cold Spring Harbor Laboratory Press, pp. 469-505, Cold Spring Harbor, New York, 2006.

Rao, N.N., Liu, S. & Kornberg, A. (1998) Inorganic Polyphosphate in *Escherichia coli*: the phosphate regulon and the stringent response. *J. Bacteriol.* **180,** 2186-2193.

Reusch, R.N., Huang, R. & Kosk-Kosicka, D. (1997) Novel components and enzymatic activities of the human erythrocyte plasma membrane calcium pump. *FEBS Lett.* **412,** 592-526.

Roinestad, F.A. & Yall, I. (1970) Volutin granules in *Zoogloea ramigera Appl. Microbiol.* **19,** 973-979.

Sandlin, R.C., Lampel, K.A., Keasler, S.P., Golberg, M.B., Stolzer, A.L. & Maurelli, A.T. (1995) Avirulence of rough mutants of *Shigella flexneri*: requirement of O antigen for correct unipolar localization of IcsA in the bacterial outer membrane. *Infect. Immu.* **63**, 229-237.

San Filippo, J. & Lambowitz, A.M. (2002) Characterization of the C-terminal DNA binding/DNA endonuclease region of a group II intron-encoded protein. *J. Mol. Biol.* **324,** 933-951.

Shih, Y.L., Kawagishi, I. & Rothfield, L. (2005) The MreB and Min cytoskeletal-like systems play independent roles in prokaryotic polar differentiation. *Mol. Microbiol.* **58,** 917-928.

Shapiro, L., McAdams, H.H., & Losick, R. (2002) Generating and exploiting polarity in bacteria. *Science* **298,** 1942-1946.

Sharp, P.A. (1991) "Five easy pieces" *Science* **254**, 663.

Simon, R., Priefer, U. & Puhler, A. (1983) A broad host range mobilization system for *in vivo* genetic engineering: transposon mutagenesis in gram negative bacteria. *Bio. Technology* **1**, 784 – 791.

Singh, N.N. & Lambowitz, A.M. (2001) Interaction of a group II intron ribonucleoprotein endonuclease with its DNA target site investigated by DNA footprinting and modification interference. *J. Mol. Biol.* **309,** 361-386.

Singh, R.N., Saldanha, R.J., D'Souza, L.M. & Lambowitz, A.M. (2002) Binding of a group II intron-encoded reverse transcriptase/maturase to its high affinity intron RNA binding site involves sequence-specific recognition and autoregulates translation. *J. Mol. Biol.* **318,** 287-303.

Slater, S., Wold, S., Lu, M., Boye, E., Skarstad, K. & Kleckner, N. (1995*) E. coli* SeqA protein binds *oriC* in two different methyl-modulated reactions appropriate to its roles in DNA replication initiation and origin sequestration. *Cell* **82,** 927-936.

Smith, D., Zhong, J., Matsuura, M., Lambowitz, A.M. & Belfort, M. (2005) Recruitment of host functions suggests a repair pathway for late steps in group II intron retrohoming. *Genes Dev.* **19**, 2477-2487.

Stevenson, G., Andrianopoulos, K., Hobbs, M. & Reeves, P.R. (1996) Organization of the *Escherichia coli* K-12 gene cluster responsible for production of the extracellular polysaccharide colanic acid. *J. Bacteriol.* **178,** 4885–4893.

Stoughton, R. B. (2005) Applications of DNA microarrays in biology. *Annu. Rev. Biochem.* **74,** 53-82.

Wank, H., San Filippo, J., Singh, R.N., Matsuura, M. & Lambowitz, A.M. (1999) A reverse transcriptase/maturase promotes splicing by binding at its own coding segment in a group II intron RNA. *Mol. Cell* **4,** 129-150.

Wheeler, D.B., Carpenter, A. E. & Sabatini, D.M. (2005) Cell microarrays and RNA interference chip away at gene function. *Nature Genetics* **37,** S25-S30.

Xiao, H., Kalman, M., Ikehara, K., Zemel, S., Glaser, G. & Cashel, M. (1991) Residual guanosine 3', 5'-bispyrophosphate synthetic activity of *relA* null mutants can be eliminated by *spoT* null mutations. *J. Biol. Chem.* **266,** 5980-5990.

Yang, J., Zimmerly, S., Perlman, P. S. & Lambowitz, A. M. (1996). Efficient integration of an intron RNA into double-stranded DNA by reverse splicing. *Nature* **381,** 332-335.

Yao, J., Zhong, J. & Lambowitz, A.M. (2005) Gene targeting using randomly inserted group II introns (targetrons) recovered from an *Escherichia coli* gene disruption library. *Nucl. Acids. Res.* **33,** 3351-3362.

Yao, J. & Lambowitz, A.M. (2007) Gene Targeting in Gram-Negative Bacteria by Use of a Mobile Group II Intron ("Targetron") Expressed from a Broad-Host-Range Vector. *Appl. Environ. Microbiol.* **73,** 2735-2743.

Young, K.D. (2006) The selective value of bacterial shape. *Microbiol. Mol. Biol. Rev.* **70,** 660-703.

Zhao, J. & Lambowitz, A.M. (2005) A bacterial group II intron-encoded reverse transcriptase localizes to cellular poles. *Proc. Natl. Acad. Sci. USA* **102,** 16133-16140.

Zhong, J., Karberg, M. & Lambowitz, A.M. (2003) Targeted and random bacterial gene disruption using a group II intron (targetron) vector containing a retrotransposition-activated selectable marker. *Nucl. Acids. Res.* **31,** 1656-1664.

Zhong, J. & Lambowitz, A.M. (2003) Group II intron mobility using nascent strands at DNA replication forks to prime reverse transcription. *EMBO J.* **22,** 4555-4565.

Zimmerly, S., Guo, H., Eskes, R., Yang, J., Perlman, P. S. & Lambowitz, A. M. (1995a). A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility. *Cell* **83,** 529-538.

Zimmerly, S., Guo, H., Eskes, R., Yang, J., Perlman, P.S., & Lambowitz, A.M. (1995b) Group II intron mobility occurs by target DNA-primed reverse transcription. *Cell* **82,** 545-554.

Vita

Junhua Zhao was born in Nanjing, Jiangsu, P.R.China, on December 13, 1976, the only daughter of Meirong Zhang and Qishan Zhao. After completing her study at the Middle School Attached to Nanjing Normal University, Nanjing, in 1995, she entered the University of Science and Technology of China, Hefei, where she received the degree of Bachelor of Science with a major in Molecular and Cellular Biology in June, 2000. After graduation, she entered the graduate program in Institute of Cellular and Molecular Biology at the University of Texas at Austin, in August, 2000. She received her degree of Doctor Philosophy in August, 2007. She coauthored the following papers:

Zhao, J. & Lambowitz, A.M. (2005) A bacterial group II intron-encoded reverse transcriptase localizes to cellular poles. *Proc. Natl. Acad. Sci. USA* **102**, 16133-16140.

Zhao, J., Niu, W., Marcotte, E.M. & Lambowitz, A.M. (2007) The intracellular localization of a group II intron-encoded reverse transcriptase and other proteins is affected by polyphosphate. In preparation.

Permanent address: Huilongqiao 8-11-104, Nanjing, Jiangsu, P.R.China 210003.

This dissertation was typed by the author.